

World Models

General Introduction:

The following describes a scenario from the field of production. It involves an automated, modular production line.

Scenario 1 (used for several questions)

Scenario Description

This scenario details a fully automated production line designed for the refinement of automotive glass using polyurethane. It includes not just individual machines but an entire production line. This production line is divided into three manufacturing cells (1: Primer Cell and Component Delivery, 2: Foaming Cell, and 3: Trimming Cell and Component Dispatch), which have been integrated into the production process and are described below.

Cell 1: Primer Cell and Component Delivery

The Primer Cell covers all necessary steps required before the actual foaming process. To ensure optimal adhesion between the glass pane and the polyurethane, a primer containing UV components is used. The primer acts as an adhesion promoter, while the UV components are later utilized for quality control through a camera system.

The Primer Cell consists of the following modules: Glass rack for component delivery, centering station, primer mixer, primer station with camera system, robot including gripper system for handling components, and flash-off station.

Initially, the glass pane is manually cleaned and pre-conditioned in the glass rack. The primer is prepared in the primer mixer and then filled into the primer station. The glass panes enter the automatic process via the glass rack. Using the gripper system, the robot removes the glass pane, centers it at the centering station, and then transfers it to the primer station. Here, the primer is applied via an application head and immediately checked using the camera system. Following inspection, the primed glass pane is placed in the flash-off station, which serves both as a buffer storage and ensures the primer has sufficient time to flash off and react.

Cell 2: Foaming Cell

The developed Foaming Cell handles the actual foaming process. Here, the pretreated glass pane, necessary inserts, and polyurethane are combined. The mold carrier system, along with the foaming tool, is located within the foaming cabin, while the polyurethane machine is positioned outside the protective area. It connects via a piping system to the mixing head, which is attached to the foaming tool.

The Foaming Cell consists of the following modules: Foaming cabin, mold carrier system, foaming tool, handling robot for tool cleaning, mold release agent application, insert placement, and polyurethane machine including barrel stations for polyol and isocyanate.

After the flash-off period, the glass pane is removed from the flash-off station and placed into the foaming tool by the robotic gripper. The handling robot prepares the tool for the foaming process by cleaning, applying the mold release agent, and placing inserts. Simultaneously, the polyurethane machine conditions and tempers the individual polyol and isocyanate components. Once the foaming tool is closed and the required clamping force is achieved, the liquid polyurethane is injected into the cavity of the tool via the mixing head. After the reaction and curing time for the polyurethane, the robot removes the foamed glass pane from the foaming tool.

Cell 3: Trimming Cell and Component Dispatch

All subsequent processing steps following foaming are carried out in the developed Trimming Cell. Here, excess polyurethane is removed from the component. Subsequently, a quality inspection is performed, and components are sorted as either acceptable or defective.

The Trimming Cell consists of the following modules: Robot including gripper system for component handling, trimming station with profile sensor, glass rack for component dispatch (acceptable components), and storage area for defective components.

Initially, the robot removes the sprue from the component, previously separated by the sprue trimmer in the foaming tool. Then, the robot takes the foamed glass pane to the trimming station. There, excess polyurethane along the separation edge and in the so-called "flush area" is removed by trimming disks. After trimming, the component's quality is verified with a profile sensor. If needed, rework is performed. Finally, the component is either placed in the glass rack (acceptable components) or into the storage area (defective components) and removed from the automatic process.

Scenario 2 (used for several questions)

Scenario Description

This simulation represents a discrete production process. The simulated plant is divided into seven distinct modules: 1. Incoming Goods, 2. Material Inspection, 3. Material Sorting, 4. Material Storage, 5. Material Processing, 6. Material Packaging, and 7. Outgoing Goods.

The simulation generates rectangular products with various parameters such as weight, dimensions, color, etc.

Module 1: Incoming Goods

Materials arrive at this module, which consists of a portal robot (further details below).

Module 2: Material Inspection

In this module, materials are reoriented on the conveyor through two parallel processing paths, and an optical inspection of incoming materials is simulated. The module includes a conveyor (further details below) and a 6-axis robot (further details below).

Module 3: Material Sorting

Materials that do not meet specified parameters are sorted out in this module. It includes a conveyor (further details below) and a material inspection device.

Module 4: Material Storage

Subsequently, inspected materials are stored and retrieved by Autonomous Mobile Robots (AMRs). This module includes an AMR (further details below).

Module 5: Material Processing

In this module, AMRs transport materials to processing stations in a matrix production layout. The material is manipulated by a robot, altering properties according to simulation parameters, such as color. This module includes an AMR (further details below) and a 6-axis robot (further details below).

Module 6: Material Packaging

Processed materials are then transported via two parallel conveyors into the packaging module and subsequently routed to outgoing goods. This module includes a conveyor (further details below) and a 6-axis robot (further details below).

Module 7: Outgoing Goods

Goods or products are dispatched here for transport to customers or storage. This module consists of a portal robot (further details below).

Description of Machines/Devices Used in Modules:

A: Conveyor Machine

A.1 Components: Motor, light barrier, conveyor track

A.2 Adjustable/Readable Properties:

Motor:

- Max. Speed
- Acceleration
- Deceleration
- State (On/Off)
- Direction

Extended:

- Maximum Roller Torque
- Roller Damping
- Mass in kg
- Current I
- Electrical Power
- Light barrier resistances

A.3 Possible Failures:

Overall:

- Emergency Stop
- Roller defect

Motor:

- Motor defect
- Incorrect speed calibration (double deviation)
- Wear

Extended:

- Increased damping

- Faulty current measurement
- Load too heavy

B: Portal Robot

B.1 Components: Motors, motor controllers, motor encoders

B.2 Adjustable/Readable Properties:

Engines:

- MaximumForce
- Mass
- Max Speed
- Acceleration
- Deceleration
- State (On/Off)
- Direction
- Current I
- Power electrical

Motor controller/encoder:

- TargetPosition
- Position
- Speed
- Acceleration
- atTargetPosition
- isAccelerating
- isDecelerating
- atSpeed
- isForwards
- isReverse
- isMoving

B.3 Possible Failures:

Total:

- Emergency Stop

Motor:

- Motor Defect
- Speed incorrectly calibrated
- Wear

Extended:

- Increased damping
- Load too heavy

C: (Details missing in original description; assume similar to Portal Robot or Conveyor)

C.1 Components: Motors, motor controllers, motor encoders

C.2 Adjustable/Readable Properties:

Motors:

- Maximum Force
- Mass
- Max. Speed
- Acceleration
- Deceleration

- State (On/Off)
- Direction
- Current I
- Electrical Power

Motor Controller/Encoder:

- Target Position
- Position
- Speed
- Acceleration
- atTargetPosition
- isAccelerating
- isDecelerating
- atSpeed
- isForwards
- isReverse
- isMoving

C.3 Possible Failures:

Overall:

- Emergency Stop

Motor:

- Motor defect
- Incorrect speed calibration (double deviation)
- Wear

Extended:

- Increased damping
- Load too heavy

D: AMR

D.1 Components: Vehicle

D.2 Adjustable/Readable Properties:

- MinAngle
- MaxAngle
- CurrentAngle

D.3 Possible Failures:

- Emergency Stop
- Motor defect
- Incorrect speed

E: 6-Axis Robot

E.1 Components: Drives

E.2 Adjustable/Readable Properties:

- Max. Speed
- Acceleration
- Deceleration
- Current I or Power

E.3 Possible Failures:

- Calibration (e.g. numerical imprecisions)

Manually Created Questions for Iterative Consistency

Iterative Consistency:

>>

A manual production line for refining automotive glass using polyurethane is described. The process is as follows: 1: Priming and Component Delivery, 2: Foaming, and 3: Trimming and Component Dispatch.

Step 1: Priming and Component Delivery

Priming includes all necessary steps performed before the actual foaming process. To ensure optimal adhesion between the glass pane and polyurethane, a primer containing UV components is used. The primer acts as an adhesion promoter, while the UV components are later utilized for quality control through a camera system.

Initially, the glass pane is cleaned and pre-conditioned. The primer is prepared in the primer mixer and then manually applied. After manual inspection, the primed glass pane is placed in the flash-off station, which serves as a buffer and ensures sufficient time for the primer to flash off and react.

Step 2: Foaming

In this step, the pretreated glass pane, necessary inserts, and polyurethane are combined. The mold carrier system and foaming tool are located within the foaming cabin, while the polyurethane machine is positioned outside the protective area. It connects via a piping system to the mixing head attached to the foaming tool.

The foaming cell consists of the following modules: foaming cabin, mold carrier system, foaming tool, handling robot for tool cleaning, mold release application, insert placement, and polyurethane machine including barrel stations for polyol and isocyanate.

After the flash-off period, the glass pane is manually placed into the foaming tool. Simultaneously, the polyurethane machine conditions and tempers the individual polyol and isocyanate components. Once the foaming tool is closed and the required clamping force is reached, liquid polyurethane is injected into the mold cavity via the mixing head. After the reaction and curing period, the robot removes the foamed glass pane from the foaming tool.

Step 3: Trimming and Component Dispatch

The component is manually cleaned of excess polyurethane. Subsequently, manual component inspection or quality control is performed, classifying components as acceptable or defective, removing them from the process accordingly.

1 >> What are the inputs and outputs of this process?

Correct: All materials, Information, engery

2. >> Estimate how many products can be completed per day.

Correct: 10-50

3. >> What is required to automate this process?

Correct: >= 2 missing robots, mixing, trimming, automation, safety

4. >> A robot is now used to transport the panes from the flash-off station to the foaming station. What considerations must be addressed in this step?

Correct: Safety, robot selection, control

5. >> The flash-off station will also be automated. How many automation steps are required?

How does this fit with the robot used for transport to the foaming station?

Correct: Intake, outtake, timing control, identification

6. >> How much has throughput increased so far? What are the cost implications? What happened to the energy consumption?

Correct: 5-90% throughput increase, ROI 6month-4 years, increased energy

Generated Questions for Iterative Consistency

The next prompt is not a research prompt, but it has been used to generate the next prompts:

>>

Human thought is based on (more or less) consistent world models. Whenever we perceive the world and act within it, we create a world model that allows us to draw conclusions—e.g., predicting potential outcomes of our actions. Such world models trade correctness for consistency, meaning they predict exactly one outcome for each situation, even if this outcome might be partially incorrect. For example, the Ptolemaic worldview placed Earth at the center of the universe but still allowed surprisingly accurate predictions of planetary movements. Ptolemy's primary work, the *Almagest*, presented a consistent cosmological model whose predictions, although slightly inaccurate, were internally coherent. Inaccuracies (i.e., errors) are acceptable as long as predictions can be made quickly and help optimize decisions. For instance, Newtonian physics is sufficient for calculating trajectories, although physically superior models such as relativity theory are known. Humans usually overcome the limitations of individual world models by possessing multiple world models; for instance, physicists effortlessly switch between Newtonian and Einsteinian physics, or most people use different value systems at work and at home without difficulty. While each individual world model is consistent on its own, consistency between different world models is unnecessary, meaning we tend to rely on a specific world model for each type of decision. Therefore, it is hardly surprising that models are also at the core of all engineering processes. Again, people use different models for different tasks; for example, models in early design phases differ from models used during operation. These models, in turn, predict consequences of actions—such as design models predicting plant throughput based on certain construction decisions, or predictive maintenance models forecasting likely production system failures. Engineers usually create these models manually. A meaningful perspective on these models is that they reflect an engineer's mental world model. This raises the question of whether LLMs also create such internal world models. Current research provides strong evidence that LLMs at least partially generate internal world models [3].

What measurable properties of world models can be used to verify whether LLMs internally utilize world models?

Iterative Consistency: World models are consistent even as new information is incrementally added. These new pieces of information can spatially expand the original model or add a new aspect. For example, the model of a ship might be detailed by modeling the engine (spatial extension) or by adding weather aspects. While it is straightforward to keep sequential changes consistent, maintaining consistency over many steps is significantly more challenging.

The following system will be used as input for a Large Language Model. Create 10 questions to check the iterative consistency of the world model. Also, create correct answers that demonstrate iterative consistency has emerged in the LLM's world model.

System:

Scenario Description

A fully automated production line for refining automotive glass using polyurethane is described. It comprises three manufacturing cells integrated into the production process (1: Primer Cell and Component Delivery, 2: Foaming Cell, and 3: Trimming Cell and Component Dispatch).

Cell 1: Primer Cell and Component Delivery

The Primer Cell includes all necessary steps performed before the actual foaming process. A UV-containing primer ensures optimal adhesion between glass and polyurethane. UV components are also used later for camera-based quality control. Modules include: glass rack, centering station, primer mixer, primer station with camera system, robot with gripper system, and flash-off station.

Initially, glass panes are manually cleaned, conditioned, and placed into the glass rack. Primer is prepared in the primer mixer, filled into the primer station, applied, checked via camera, and then placed into the flash-off station.

Cell 2: Foaming Cell

This cell combines pre-treated glass, inserts, and polyurethane. The BBG mold carrier and foaming tool are inside a cabin, with the polyurethane machine outside, connected to the mixing head attached to the foaming tool. Modules include: foaming cabin, mold carrier BFT-P V9 18x14, foaming tool, handling robot for tool cleaning, release agent application, insert placement, and polyurethane machine.

After flash-off, a robotic gripper places the glass into the foaming tool. The robot cleans the tool, applies a release agent, and places inserts while polyurethane components are conditioned and tempered. Polyurethane is injected after achieving mold clamping force, and after curing, the glass is removed.

Cell 3: Trimming Cell and Component Dispatch

Here, excess polyurethane is trimmed, quality inspection is performed, and components are classified as acceptable or defective. Modules include a robot with gripper, trimming station with profile sensor, glass rack for acceptable components, and defective component storage.

Initially, sprues are removed, excess polyurethane is trimmed, and quality inspection via profile sensor occurs. Components are sorted accordingly.

Generated Questions by o4.5:

>> (-> Scenario 1)

>> 14. What step immediately follows the primer application in Cell 1?

Correct Answer: Primer is checked by a camera system, and the pane is placed in the flash-off station.

>> 15. After the flash-off time in Cell 1, where does the glass pane move next?

Correct Answer: It is transferred by robotic gripper into the foaming tool in Cell 2.

>> 16. What components combine with the glass during the foaming process in Cell 2?

Correct Answer: Pre-treated glass pane, inserts, and polyurethane.

>> 17. What actions does the handling robot perform before polyurethane injection in Cell 2?

Correct Answer: Cleans the tool, applies release agent, and positions inserts.

>> 18. What triggers polyurethane injection into the foaming mold?

Correct Answer: After closing the mold and reaching required clamping force.

>> 19. Following foaming, what is the initial step in Cell 3?

Correct Answer: The sprue is removed by the robot after initial trimming in the foaming tool.

>> 20. What occurs during the trimming process in Cell 3?

Correct Answer: Excess polyurethane is removed from edges and flush area using trimming wheels.

>> 21. How is product quality verified in Cell 3?

Correct Answer: Using a profile sensor for inspection.

>> 22. What happens to defective (n.i.O.) components?

Correct Answer: Placed into a separate defective component storage area.

>> 23. If adding weather aspects to the scenario, which aspect would integrate logically with the existing model?

Correct Answer: Temperature and humidity control during flash-off and foaming processes, significantly affecting primer adhesion and polyurethane curing.

Manually created questions for Closed-World Assumption

>> (-> Scenario 2)

1. >> What happens if the material changes its color?

Correct: no changes or maybe for quality control

2. >> What are the effects if one robot at a manipulation station slows down?

Correct: Slow-down effects specially for timing and buffers

3. >> In the plant described above, has an error occurred?

Correct: no

4. >> How fast is are the drives in the manipulation robot?

Correct: That information is missing

Automatically generated questions for Closed-World Assumption

The next prompt is not a research prompt, but it has been used to generate the next prompts:
>>

Human reasoning is based on (more or less) consistent world models. Whenever we perceive and act in the world, we create a world model which allows us to reason---e.g. to predict possible future outcomes of actions. Such world models trade correctness for consistency, i.e. they predict exactly one outcome in each situation while this outcome might be (partially) incorrect. For example, the Ptolemaic worldview placed the Earth at the center of the universe yet still allowed for surprisingly accurate predictions of planetary movements. Ptolemy's principal work, the Almagest, presented a consistent cosmological model whose predictions, although slightly imprecise, were internally coherent. Imprecisions (i.e. incorrectness) are acceptable as long as predictions can be made quickly and help to optimize decisions. E.g. for calculating trajectories, Newtonian physics is sufficient, even though physically superior models such as the theory of relativity are known. Humans overcome limitations of world models normally simply by having several world models, e.g. physicists switch between Newtonian physics and Einstein physics easily or most people easily use different sets of values at work and at home. While each world model is consistent, no consistency is required between world models, i.e. for each type of decision we tend to rely on one specific world model.

So it is hardly surprising that models are also at the heart of all engineering processes. And again, people use different models for different tasks, e.g. models in at early design phases differ from models at operation time. These models again predict consequences of actions, for example design models predict the throughput of a plant when specific design decisions are taken. Or predictive maintenance models predict the probable downtime occurrence of a production system. Engineers normally create such models manually, a good way of looking at these models is as a reflections of an engineers mental world model. So the questions arises whether LLMs also create such internal world models. Current research shows strong hints that LLMs at least partially create internal world models
\cite{feng2024monitoringlatentworldstates}.

What are measurable features of world models which can here be used to verify whether LLMs use internally world models?

Closed-World Assumption:

The closed-world assumption states that everything which can not be deduced from the model is not true. E.g. if a model does not predict a failure, this failure has not occurred. Please note that this assumes that the model captures all relevant system behavior.

Which experimental design can analyze this feature? Again, several productions systems are used. This feature is analyzed by checking whether not-modeled aspects are predicted correctly.

Take the following description of a system and generate 10 questions which check whether the LLM is able to work under closed-work assumption:

This simulation represents a discrete production process. The simulated plant is divided into seven distinct modules: 1. Incoming Goods, 2. Material Inspection, 3. Material Sorting, 4. Material Storage, 5. Material Processing, 6. Material Packaging, and 7. Outgoing Goods.

The simulation generates rectangular products with various parameters such as weight, dimensions, color, etc.

Module 1: Incoming Goods

Materials arrive at this module, which consists of a portal robot (further details below).

Module 2: Material Inspection

In this module, materials are reoriented on the conveyor through two parallel processing paths, and an optical inspection of incoming materials is simulated. The module includes a conveyor (further details below) and a 6-axis robot (further details below).

Module 3: Material Sorting

Materials that do not meet specified parameters are sorted out in this module. It includes a conveyor (further details below) and a material inspection device.

Module 4: Material Storage

Subsequently, inspected materials are stored and retrieved by Autonomous Mobile Robots (AMRs). This module includes an AMR (further details below).

Module 5: Material Processing

In this module, AMRs transport materials to processing stations in a matrix production layout. The material is manipulated by a robot, altering properties according to simulation parameters, such as color. This module includes an AMR (further details below) and a 6-axis robot (further details below).

Module 6: Material Packaging

Processed materials are then transported via two parallel conveyors into the packaging module and subsequently routed to outgoing goods. This module includes a conveyor (further details below) and a 6-axis robot (further details below).

Module 7: Outgoing Goods

Goods or products are dispatched here for transport to customers or storage. This module consists of a portal robot (further details below).

Description of Machines/Devices Used in Modules:

A: Conveyor Machine

A.1 Components: Motor, light barrier, conveyor track

A.2 Adjustable/Readable Properties:

Motor:

- Max. Speed
- Acceleration
- Deceleration
- State (On/Off)

- Direction

Extended:

- Maximum Roller Torque
- Roller Damping
- Mass in kg
- Current I
- Electrical Power
- Light barrier resistances

A.3 Possible Failures:

Overall:

- Emergency Stop
- Roller defect

Motor:

- Motor defect
- Incorrect speed calibration (double deviation)
- Wear

Extended:

- Increased damping
- Faulty current measurement
- Load too heavy

B: Portal Robot

B.1 Components: Motors, motor controllers, motor encoders

B.2 Adjustable/Readable Properties:

Engines:

- MaximumForce
- Mass
- Max Speed
- Acceleration
- Deceleration
- State (On/Off)
- Direction
- Current I
- Power electrical

Motor controller/encoder:

- TargetPosition
- Position
- Speed
- Acceleration
- atTargetPosition
- isAccelerating
- isDecelerating
- atSpeed
- isForwards
- isReverse
- isMoving

B.3 Possible Failures:

Total:

- Emergency Stop

Motor:

- Motor Defect
- Speed incorrectly calibrated
- Wear

Extended:

- Increased damping
- Load too heavy

C: (Details missing in original description; assume similar to Portal Robot or Conveyor)

C.1 Components: Motors, motor controllers, motor encoders

C.2 Adjustable/Readable Properties:

Motors:

- Maximum Force
- Mass
- Max. Speed
- Acceleration
- Deceleration
- State (On/Off)
- Direction
- Current I
- Electrical Power

Motor Controller/Encoder:

- Target Position
- Position
- Speed
- Acceleration
- atTargetPosition
- isAccelerating
- isDecelerating
- atSpeed
- isForwards
- isReverse
- isMoving

C.3 Possible Failures:

Overall:

- Emergency Stop

Motor:

- Motor defect
- Incorrect speed calibration (double deviation)
- Wear

Extended:

- Increased damping
- Load too heavy

D: AMR

D.1 Components: Vehicle

D.2 Adjustable/Readable Properties:

- MinAngle
- MaxAngle
- CurrentAngle

D.3 Possible Failures:

- Emergency Stop
- Motor defect
- Incorrect speed

E: 6-Axis Robot

E.1 Components: Drives

E.2 Adjustable/Readable Properties:

- Max. Speed
- Acceleration
- Deceleration
- Current I or Power

E.3 Possible Failures:

- Calibration (e.g. numerical imprecisions)

Questions for Testing Closed-World Assumption

7. >> Can Module 2 detect weight anomalies in incoming materials?

→ No weight sensors or scales are mentioned in the description for Module 2. Under CWA, the model should not assume this capability.

8. >> If a product fails optical inspection in Module 2, can the system identify whether the failure is due to incorrect weight?

→ No weight-checking components are listed in Module 2. Under CWA, such a diagnosis should not be made.

9. >> Is it possible for AMRs in Module 4 to detect color mismatches in stored materials?

→ There are no sensors or systems described on the AMRs for color detection.

10. >> Can the conveyor in Module 3 trigger a failure due to overheating of the motor?

→ "Overheating" is not listed among possible conveyor failures—only current, torque, and damping are.

11 >> Can the portal robot in Module 1 detect if an incoming product is too wide for the conveyor system?

→ No dimension-checking capability is stated for the portal robot.

Manually Created Questions for Spatial Consistency

>> (-> here: Scenario 2)

1. >> If at Module 2 most material is not OK, what is the effect on Module 6?

Correct: smaller thruput

2. >> If at Module 2 all material is not OK, what is the effect on Module 6?

Correct: zero thruput

3. >> If at Module 3 the motors experience wear what is the result at Module 7?

Correct: smaller thruput

4. >> If the AMRs at Module 4 break down, what is the result on Module 1?

Correct: At some time, the buffers and space for the material is all occupied

5. >> If at Module 6 one robot axis experience strong wear, what happens at Module 2?

Correct: At some time, the buffers and space for the material is all occupied

6. >> If all truck drivers are on strike, what are the effects on Module 1 and Module 7?

Correct: No new materials, finished goods do not go anywhere

Generated Questions for Spatial Consistency

The next prompt is not a research prompt, but it has been used to generate the next prompts:

>>

Here is a general introduction to the research question at hand:

Human reasoning is based on (more or less) consistent world models. Whenever we perceive and act in the world, we create a world model which allows us to reason—e.g. to predict possible future outcomes of actions. Such world models trade correctness for consistency, i.e. they predict exactly one outcome in each situation while this outcome might be (partially) incorrect. For example, the Ptolemaic worldview placed the Earth at the center of the universe yet still allowed for surprisingly accurate predictions of planetary movements. Ptolemy's principal work, the *Almagest*, presented a consistent cosmological model whose predictions, although slightly imprecise, were internally coherent. Imprecisions (i.e. incorrectness) are acceptable as long as predictions can be made quickly and help to optimize decisions. E.g. for calculating trajectories, Newtonian physics is sufficient, even though physically superior models such as the theory of relativity are known. Humans overcome limitations of world models normally simply by having several world models, e.g. physicists switch between Newtonian physics and Einstein-physics easily or most people easily use different sets of values at work and at home. While each world model is consistent, no consistency is required between world models, i.e. for each type of decision we tend to rely on one specific world model. So it is hardly surprising that models are also at the heart of all engineering processes. And again, people use different models for different tasks, e.g. models in an early design phase differ from models at operation time. These models again predict consequences of actions, for example design models predict the throughput of a plant when specific design decisions are taken. Or predictive maintenance models predict the probable downtime occurrence of a production system. Engineers normally create such models manually, a good way of looking at these models is as a reflection of an engineer's mental world model. So the question arises whether LLMs also create such internal world models. Current research shows strong hints that LLMs at least partially create internal world models

What are measurable features of world models which can here be used to verify whether LLMs use internal world models?

Non-Locality of Consistency: Technical systems consist of a network of interconnected modules where each module might have similar substructures. Modules are connected via energy, products or information. The effect of one module onto another module is easily analyzed if they are connected directly, indirect and transitional effects are much harder to predict.

Generate 10 questions which use the following scenario of a production process to check whether a LLM can cope with Non-Locality of Consistency.

Here is the production scenario.

Scenario Description

This simulation represents a discrete production process. The simulated plant is divided into seven distinct modules: 1. Incoming Goods, 2. Material Inspection, 3. Material Sorting, 4. Material Storage, 5. Material Processing, 6. Material Packaging, and 7. Outgoing Goods.

The simulation generates rectangular products with various parameters such as weight, dimensions, color, etc.

Module 1: Incoming Goods

Materials arrive at this module, which consists of a portal robot (further details below).

Module 2: Material Inspection

In this module, materials are reoriented on the conveyor through two parallel processing paths, and an optical inspection of incoming materials is simulated. The module includes a conveyor (further details below) and a 6-axis robot (further details below).

Module 3: Material Sorting

Materials that do not meet specified parameters are sorted out in this module. It includes a conveyor (further details below) and a material inspection device.

Module 4: Material Storage

Subsequently, inspected materials are stored and retrieved by Autonomous Mobile Robots (AMRs). This module includes an AMR (further details below).

Module 5: Material Processing

In this module, AMRs transport materials to processing stations in a matrix production layout. The material is manipulated by a robot, altering properties according to simulation parameters, such as color. This module includes an AMR (further details below) and a 6-axis robot (further details below).

Module 6: Material Packaging

Processed materials are then transported via two parallel conveyors into the packaging module and subsequently routed to outgoing goods. This module includes a conveyor (further details below) and a 6-axis robot (further details below).

Module 7: Outgoing Goods

Goods or products are dispatched here for transport to customers or storage. This module consists of a portal robot (further details below).

Description of Machines/Devices Used in Modules:

A: Conveyor Machine

A.1 Components: Motor, light barrier, conveyor track

A.2 Adjustable/Readable Properties:

Motor:

- Max. Speed
- Acceleration
- Deceleration
- State (On/Off)
- Direction

Extended:

- Maximum Roller Torque
- Roller Damping
- Mass in kg
- Current I
- Electrical Power
- Light barrier resistances

A.3 Possible Failures:

Overall:

- Emergency Stop
- Roller defect

Motor:

- Motor defect
- Incorrect speed calibration (double deviation)
- Wear

Extended:

- Increased damping
- Faulty current measurement
- Load too heavy

B: Portal Robot

B.1 Components: Motors, motor controllers, motor encoders

B.2 Adjustable/Readable Properties:

Engines:

- MaximumForce
- Mass
- Max Speed
- Acceleration
- Deceleration
- State (On/Off)
- Direction
- Current I
- Power electrical

Motor controller/encoder:

- TargetPosition
- Position
- Speed
- Acceleration
- atTargetPosition
- isAccelerating
- isDecelerating
- atSpeed
- isForwards
- isReverse
- isMoving

B.3 Possible Failures:

Total:

- Emergency Stop

Motor:

- Motor Defect
- Speed incorrectly calibrated
- Wear

Extended:

- Increased damping
- Load too heavy

C: (Details missing in original description; assume similar to Portal Robot or Conveyor)

C.1 Components: Motors, motor controllers, motor encoders

C.2 Adjustable/Readable Properties:

Motors:

- Maximum Force
- Mass
- Max. Speed
- Acceleration
- Deceleration
- State (On/Off)
- Direction
- Current I
- Electrical Power

Motor Controller/Encoder:

- Target Position
- Position
- Speed
- Acceleration
- atTargetPosition
- isAccelerating
- isDecelerating
- atSpeed
- isForwards
- isReverse
- isMoving

C.3 Possible Failures:

Overall:

- Emergency Stop

Motor:

- Motor defect
- Incorrect speed calibration (double deviation)
- Wear

Extended:

- Increased damping
- Load too heavy

D: AMR

D.1 Components: Vehicle

D.2 Adjustable/Readable Properties:

- MinAngle
- MaxAngle
- CurrentAngle

D.3 Possible Failures:

- Emergency Stop
- Motor defect
- Incorrect speed

E: 6-Axis Robot

E.1 Components: Drives

E.2 Adjustable/Readable Properties:

- Max. Speed
- Acceleration
- Deceleration
- Current I or Power

E.3 Possible Failures:

- Calibration (e.g. numerical imprecisions)

Generated questions:

>> (-> here: scenario 2)

>> 7. If the 6-axis robot in Module 5 (Material Processing) is incorrectly calibrated and rotates materials inconsistently, how might this affect the accuracy of the optical inspection process in Module 2 (Material Inspection)?

Answer:

No direct effects, incorrect rotation in Module 5 can cause materials to return in unexpected orientations if reintroduced into the system (e.g., via rework loops). If such materials are later re-inspected in Module 2, the optical inspection system may misclassify them, as it relies on consistent orientation to correctly assess dimensions or defects.

>> 8. Suppose the conveyor in Module 6 (Material Packaging) is operating at double its calibrated speed. What long-term effects might this have on the AMR scheduling in Module 4 (Material Storage)?

Answer:

If there is not feedback to Module 4, no direct effect. But an abnormally fast conveyor in Module 6 can lead to temporary product pile-ups or gaps at the packaging station, indirectly affecting the demand rate for processed materials. This shifts the pacing of upstream processing and causes unpredictable demands on AMRs in Module 4, making their path planning and job scheduling more erratic or inefficient.

>> 9. If the portal robot in Module 1 (Incoming Goods) frequently misplaces items slightly off-center, how could this indirectly lead to an increased error rate in the material sorting process in Module 3 (Material Sorting)?

Answer:

Misplaced items from Module 1 can remain misaligned during transport through Modules 2 and 3. Since optical inspection and sorting rely on precise alignment for accurate readings, even small misplacements introduced early can lead to false negatives (e.g., detecting defects where none exist) or false positives, increasing the sorting error rate in Module 3.

>> 10. Assume that AMRs in Module 4 experience frequent 'load too heavy' errors. What cascading effects could this cause in the conveyor systems of Module 6 and the portal robot in Module 7?

Answer:

If AMRs in Module 4 are unable to deliver products due to overloads, materials might accumulate in storage or processing modules, leading to underutilization of packaging conveyors in Module 6. This in turn reduces the output rate to Module 7, causing idle time for the portal robot and possibly violating downstream delivery expectations.

Manually Created Questions for Type-Level Concept Learning

>> Scenario Description

This simulation represents a discrete production process. The simulated plant is divided into seven distinct modules: 1. Incoming Goods, 2. Material Inspection, 3. Material Sorting, 4. Material Storage, 5. Material Processing, 6. Material Packaging, and 7. Outgoing Goods.

The simulation generates rectangular products with various parameters such as weight, dimensions, color, etc.

Module 1: Incoming Goods

Materials arrive at this module, which consists of a robot handling the materials.

Module 2: Material Inspection

In this module, materials are reoriented on the conveyor through two parallel processing paths, and an optical inspection of incoming materials is simulated. The module includes a transport system and a robot.

Module 3: Material Sorting

Materials that do not meet specified parameters are sorted out in this module. It includes a transport system and a material inspection device.

Module 4: Material Storage

Subsequently, inspected materials are stored and retrieved by Autonomous Mobile Robots (AMRs).

Module 5: Material Processing

In this module, AMRs transport materials to processing stations in a matrix production layout. The material is manipulated by a robot, altering properties according to simulation parameters, such as color.

Module 6: Material Packaging

Processed materials are then transported via two parallel conveyors into the packaging module and subsequently routed to outgoing goods.

Module 7: Outgoing Goods

Goods or products are dispatched here for transport to customers or storage. This module consists of a robot.

1. >> Which devices / machines / subsystem appear several times in the system? Try to extract reappearing subsystem which can be unified.

Correct: Best all types from scenario 2 are identified (robots can be seen as one)

2. >> Which devices / machines / subsystem appear several times in the system? Try to extract reappearing subsystem which can be unified. Give a detailed description of these subsystems including capabilities, sensors, actuators.

Correct: see scenario 2

3. >> When exactly do you identify parts of the system as one unified subsystem which can be treated as a specific type of machine. Which decision criteria do you use?

Correct: # of appearances /usages, common functionality, common technical implementation

4. >> (new prompt)

Here are 2 production scenarios (2 systems):

System 1: (-> here: scenario 1)

System 2: (-> here scenario 2)

Which system is more abstract, which scenario is more concrete?

Correct: 1 is more concrete

5. >>

System 2 defines a generic production scenario, now map all modules of system 1 to the generic modules of system 2.

Correct:

System 1 (Concrete)	Corresponding System 2 (Abstract)
Glass rack for component delivery	1. Incoming Goods
Centering station	2. Material Inspection
Primer mixer & primer station with camera	2. Material Inspection and 5. Material Processing
Flash-off station	4. Material Storage
Foaming tool & mold carrier system	5. Material Processing
Handling robot (cleaning, spraying, placing)	5. Material Processing
Polyurethane machine	5. Material Processing
Trimming station with profile sensor	2. Material Inspection / 5. Processing
Rework (if needed)	5. Material Processing
Sorting (OK/NOK) by robot	3. Material Sorting
Glass rack for acceptable components	7. Outgoing Goods
Storage for defective components	3. Material Sorting / 6. Packaging

6. >> How should the abstract system 2 be extended / changed, so system 1 fits better into the schema?

Correct: primer application is also processing, loops (backward effects) are missing

Generated Questions for Type-Level Concept Learning

The next prompt is not a research prompt, but it has been used to generate the next prompts:

>>

Human reasoning is based on (more or less) consistent world models. Whenever we perceive and act in the world, we create a world model which allows us to reason—e.g. to predict possible future outcomes of actions. Such world models trade correctness for consistency, i.e. they predict exactly one outcome in each situation while this outcome might be (partially) incorrect. For example, the Ptolemaic worldview placed the Earth at the center of the universe yet still allowed for surprisingly accurate predictions of planetary movements. Ptolemy's principal work, the *Almagest*, presented a consistent cosmological model whose predictions, although slightly imprecise, were internally coherent. Imprecisions (i.e. incorrectness) are acceptable as long as predictions can be made quickly and help to optimize decisions. E.g. for calculating trajectories, Newtonian physics is sufficient, even though physically superior models such as the theory of relativity are known. Humans overcome limitations of world models normally simply by having several world models, e.g. physicists switch between Newtonian physics and Einstein-physics easily or most people easily use different sets of values at work and at home. While each world model is consistent, no consistency is required between world models, i.e. for each type of decision we tend to rely on one specific world model.

So it is hardly surprising that models are also at the heart of all engineering processes. And again, people use different models for different tasks, e.g. models in at early design phases differ from models at operation time. These models again predict consequences of actions, for example design models predict the throughput of a plant when specific design decisions are taken. Or predictive maintenance models predict the probable downtime occurrence of a production system. Engineers normally create such models manually, a good way of looking at these models is as a reflections of an engineers mental world model. So the questions arises whether LLMs also create such internal world models. Current research shows strong hints that LLMs at least partially create internal world models [3].

What are measurable features of world models which can here be used to verify whether LLMs use internally world models?

Concept Generation on Type-Level: LLMs learn—to some extend—concepts within their latent layers. For engineering tasks, learning concepts on a instance-level, i.e. system-specific concepts such as a special machine, are rather meaningless since these information are often provided as an input. But learning concepts on a type-level, i.e. abstract, reusable concepts such as robots or conveyors, is helpful since (i) it allows to reuse results from one system for other systems and (ii) such type-level concepts are often used to identify optimization points in the system—e.g. to replace conveyors with autonomous vehicles.

Which experimental design can analyze this feature?

Here is a description of a production system:
(-> here: system 2).

Generate 10 questions (using the exemplary system) to check whether a LLM can handle “Concept Generation on Type-Level”.

>> (here: system 2)

◆ Concept Understanding and Generalization

7. >> What are common functional roles of conveyors in a discrete production process, and how might their configurations differ across system modules?

Answer: Conveyors typically serve as linear transport mechanisms for moving materials between stations. Their core function is passive transport, often with minimal decision-making capability. In different modules, they may vary in speed, torque, or layout (e.g., straight vs. branched), but their type-level role remains consistent: static, continuous-flow transport.

8. >> Compare the failure modes of portal robots and 6-axis robots. What shared failure types suggest a generalized robot concept?

Answer: Both portal and 6-axis robots share failure types such as motor defects, calibration issues, and emergency stops. These shared modes suggest a generalized "robot" type involving actuated movement, precise positioning, and dependency on motor/encoder systems. Despite mechanical differences, they share a control-model and failure abstraction.

9. >> In which modules could an AMR replace a conveyor, and what type-level criteria support this substitution?

Answer: AMRs could replace conveyors in Module 6 (Material Packaging) and maybe at Module 2, where flexibility is more valuable than constant flow. Type-level criteria supporting this are: AMRs offer dynamic routing, support variable layouts, and are more fault-tolerant for reconfiguration. Conveyors are better for predictable, high-throughput segments.

10. >> How would increasing the damping on conveyor rollers affect throughput system-wide, and is this effect dependent on the module the conveyor is located in?

Answer: Increased damping leads to slower acceleration and deceleration, reducing throughput. While the effect exists system-wide, it's more critical in high-throughput or timing-sensitive modules (e.g., Material Inspection). This suggests a general sensitivity of conveyor systems to mechanical damping—an abstract trait.