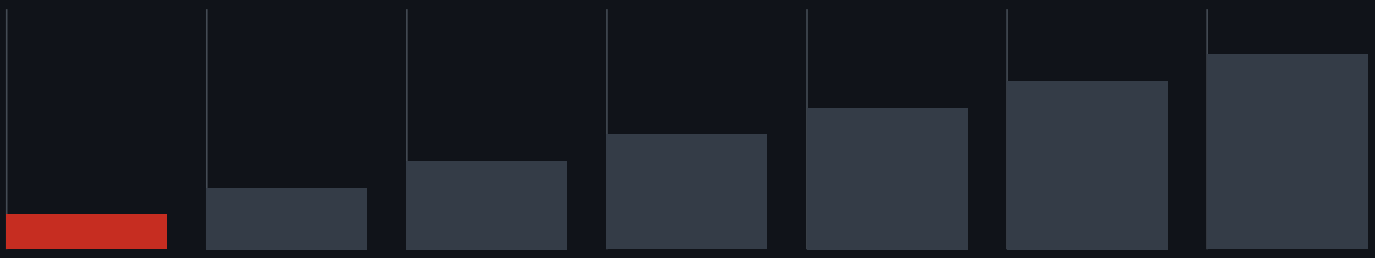
UNDERSTAND INTELLIGENCE.

Inside AI.

7 Concepts Everyone Should Understand

BY ALDIYAR SEMEDYAROV



A visual field guide by Aldiyar Semedyarov. Version 1.0, 28 October 2026.

Understand seven essential AI concepts, see their limits, and know what to check before trusting an AI answer. Allow 20-30 minutes, including the short checks. You need simple arithmetic, no programming and no paid AI account.

Follow one fictional engineering task: Sam asks an assistant to summarize a note about a mobile robot. The robot slows near a person. The note reports 12 successful stops in 15 trials. Can Sam trust the summary, its calculation or a claim that the robot is safe?

By the end, explain how text becomes numbers, why prediction differs from verification, and why a convincing chatbot answer does not establish physical competence. Keep page 13 for your next AI-assisted task.

CHECK YOUR UNDERSTANDING

What will you practice judging?

Answer: An AI output against its task, evidence and consequences.

02 / INSIDE AI

One task, seven connected ideas

Locate each concept in the running task and choose a reading route.

Sam supplies this fictional note: "The robot slowed near the person. It stopped successfully in 12 of 15 trials." The assistant must summarize it and calculate the fraction. Later, Sam asks whether the robot could operate around people. These are different tasks, needing different evidence.

01 Tokenization and embeddings: turn the note into token IDs and numerical representations (pages 3-4).

02 Networks: combine numbers using learned parameters; distinguish learning from use (pages 5-6).

03 Transformers: let representations exchange information through attention (page 7).

04 Language modeling: generate a continuation one token at a time (page 8).

05 Reasoning and limits: assess multistep behavior and its boundaries (pages 9-10).

06 Hallucination and reliability: inspect evidence behind a convincing answer (page 11).

07 Robotics: connect sensing, decisions and physical feedback (page 12).

Take the intuition route by reading the prose, diagrams and checks. For the arithmetic route, also work through pages 5 and 7 with a calculator or paper. Both routes finish with the same audit.

This map concentrates on text-generating transformer systems. AI also includes other architectures and tasks. The guide does not teach model implementation, establish consciousness, qualify you to deploy robots or provide professional safety approval.

CHECK YOUR UNDERSTANDING

Does a good summary answer the physical-safety question?

Answer: No. Summarization evidence and physical-system evidence differ.

03 / INSIDE AI

Tokens are pieces, IDs are labels

Distinguish a token, a word and an arbitrary token ID.

A tokenizer converts text into a sequence of pieces and their vocabulary IDs. A piece need not be a whole word: subwords and punctuation can be separate tokens. The exact split depends on the tokenizer. [T03]

Use this invented tokenizer for "The robot slowed." It makes five pieces: [The], [robot], [slow], [ed], [.]. The leading spaces belong to the pieces. Our made-up IDs are [11, 42, 73, 8, 5]. Concatenating the pieces restores the sentence. These are not outputs from a named production tokenizer.

The ID 42 selects a vocabulary entry. It does not say that " robot" is twice as meaningful as ID 21, and subtracting two IDs tells you nothing useful about their meanings. A model needs numerical representations to compute with the entries; page 4 adds those.

Think of an ID as a catalogue number. The analogy explains identification only: a catalogue number does not describe a book, and a token ID is not a semantic description.

THE MECHANISM

The robot slowed.

INVENTED TOKENIZER

Piece	ID
The	11
[space]robot	42
[space]slow	73
ed	8
.	5

[space] means one literal space.

D01. Invented pieces and IDs; not a production tokenizer. Pieces concatenate exactly to the sentence.

SOURCES T03 / LINKED IN THE TEXT COMPANION

CHECK YOUR UNDERSTANDING

Why are there five tokens but only three words in the toy sentence?

Answer: The toy splits "slowed" into two pieces and counts the full stop separately.

04 / INSIDE AI

Representations let a network compute

Separate an ID, its embedding and a contextual hidden state.

An embedding lookup uses each ID to select a vector: an ordered list of numbers. In an ordinary learned input-embedding table, the same ID selects the same row while the table is fixed. The ID is an address; the vector supplies values for computation. [T04]

Let ID 42 select the invented vector [0.2, -0.4, 0.7]. No coordinate is labeled "robotness." Real representations generally have many more coordinates; interpretation depends on the model and training.

The network then transforms input representations using position and context. A contextual hidden state for "robot" in a sentence about a factory can differ from its state in a sentence about a toy. That changing state is not the same object as the original fixed lookup row. [T01]

Our toy 2D projection keeps the first two coordinates and drops the third. Two different 3D vectors can then land on the same point. Projection distances cannot establish universal meaning.

For Sam, a representation influenced by "person" is not a sensor reading of an actual person. Describing the world and observing it remain different things.

THE MECHANISM

ID IS AN INDEX

42 -> A = [0.2, -0.4, 0.7]

B = [0.2, -0.4, -0.6]

TOY PROJECTION: KEEP x AND y



Same 2D point: [0.2, -0.4]

Different vectors can overlap.

Hidden states also depend on context.

D01b. Invented vectors and projection; the third coordinate is lost. This is not a measured semantic map.

SOURCES T04 · T01 / LINKED IN THE TEXT COMPANION

CHECK YOUR UNDERSTANDING

If context changes a hidden state, must the embedding table have been retrained?

Answer: No. Computation can change contextual states while the lookup table stays fixed.

05 / INSIDE AI

A small network calculation

Reproduce a weighted sum, bias and ReLU output.

A basic artificial neuron multiplies inputs by weights, adds a bias, then applies a function. Weights and biases are parameters. A nonlinear function allows stacked layers to express relationships beyond a single affine transformation. This is a mathematical building block, not a realistic biological neuron. [T13]

Try this independently authored example. Inputs are x1 = 2 and x2 = 1. Weights are w1 = 0.5 and w2 = -1. The bias b = 0.5. First calculate the pre-activation z = w1*x1 + w2*x2 + b.

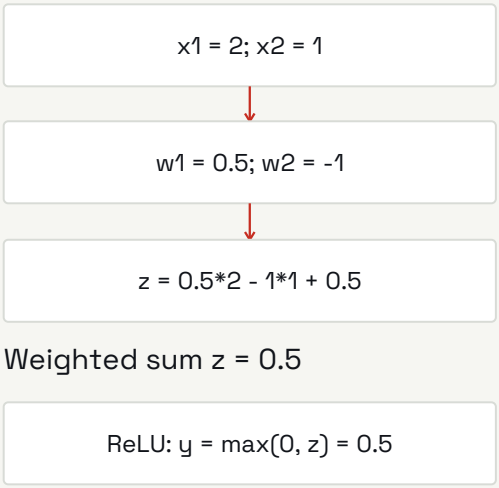
z = 0.5*2 + (-1)*1 + 0.5 = 0.5

Now use ReLU, which returns the larger of zero and its input: y = max(0, z). The result is y = 0.5. If z had been -0.5, ReLU would have returned zero. The result here is a toy activation, not a probability or a robot-safety score.

Change x2 to 2: z becomes -0.5 and y becomes zero. Large networks combine many operations, often with other nonlinear functions.

The inputs are unnamed features, not measured distance or speed. This tiny circuit is not a working language model or controller.

THE MECHANISM



D02. Original two-input example. Bias is 0.5; ReLU clips negative pre-activations to zero. All quantities are specified.

SOURCES T13 / LINKED IN THE TEXT COMPANION

CHECK YOUR UNDERSTANDING

With x1 = 4 and x2 = 1, what is y?

Answer: z = 2 - 1 + 0.5 = 1.5, so y = 1.5.

06 / INSIDE AI

Learning is different from using

Identify what changes in training and ordinary inference.

During training, a system produces outputs on examples, calculates a loss against a training objective, and updates parameters using an optimization procedure. Loss measures mismatch under that objective; it is not a direct score for truth, usefulness or safety. Evaluation uses separate examples to check how well behavior carries over. [T05]

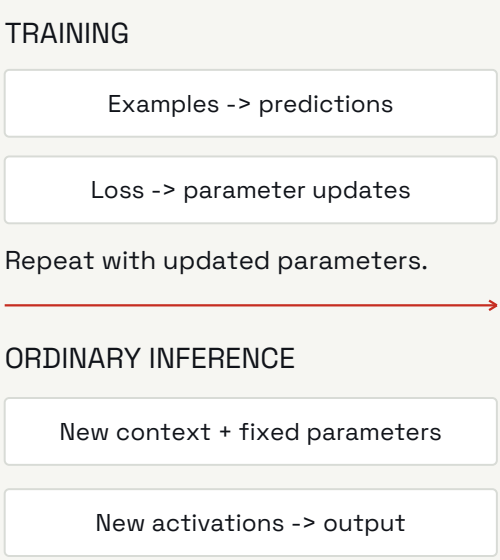
During ordinary inference, the trained parameters stay fixed while new inputs produce new activations and outputs. Sam can add a note or correct a sentence, changing the available context and the answer without changing the model's weights. Brown and colleagues demonstrated few-shot tasks through supplied text without gradient updates. [T15]

Think of adjusting a recipe, then cooking from it. The analogy stops at adjustment: a network performs numerical optimization. Some deployed systems adapt parameters or maintain separate memory; fixed parameters describe ordinary inference, not every system.

If a model memorizes examples but performs poorly on unfamiliar ones, the training result has not generalized as intended. Test on cases that resemble the real task, including failures; a few familiar successes are weak evidence. [T05]

For Sam, training quality, available evidence and answer verification are separate questions.

THE MECHANISM



D03. A changing output does not prove a parameter update. Adaptive deployed systems are outside this fixed-parameter example.

SOURCES T05 · T15 / LINKED IN THE TEXT COMPANION

CHECK YOUR UNDERSTANDING

Sam pastes a correction into a chat. Does that alone prove a weight update?

Answer: No. A changed context can change the output without retraining.

07 / INSIDE AI

Attention combines information

Calculate one weighted output and explain what it cannot establish.

In a transformer, attention computes weights from query-key compatibility and combines value vectors. Queries, keys and values are numerical projections of representations, not questions, dictionary entries or human judgments. Multiple heads and layers participate in the full network. [T01]

Take ”The robot slowed near the person.” At the final full stop, show only two allowed positions: ”robot” and ”person.” Invent normalized weights 0.25 and 0.75, with scalar values 2 and 6. The weighted output is:

0.25*2 + 0.75*6 = 5

The weights add to one. If the weights change to 0.5 and 0.5 while the values stay fixed, the output becomes 4. This is a mixing calculation. The numbers are not measurements from a real model, and 0.75 is not a probability that ”person” is the correct explanation.

A decoder’s causal mask blocks attention to future positions. Our simplification omits other positions, score calculation and the rest of the network. [T01]

Attention weights alone can be misleading as explanations, as research on attention-based NLP systems shows. The diagram does not establish why a whole model answered as it did. [T06]

THE MECHANISM

TOY: TWO ALLOWED POSITIONS

Query position: final full stop

robot: 0.25 * 2 = 0.5

person: 0.75 * 6 = 4.5

Weights: 0.25 + 0.75 = 1

Output: 0.5 + 4.5 = 5

Weights alone do not establish why the model answered.

D04. Invented normalized weights and scalar values. Only a weighted mixture is shown; it is not a real model trace.

SOURCES T01 · T06 / LINKED IN THE TEXT COMPANION

CHECK YOUR UNDERSTANDING

If the larger weight moves from value 6 to value 2, does the output rise or fall?

Answer: It falls: 0.75*2 + 0.25*6 = 3.

08 / INSIDE AI

Prediction becomes a sequence

Explain autoregressive generation and separate probability from verification.

An autoregressive language model assigns probabilities to possible next tokens given the available preceding context. After selecting a token, generation adds it to the context and repeats. This can produce a long, coherent answer through repeated next-token steps. [T15]

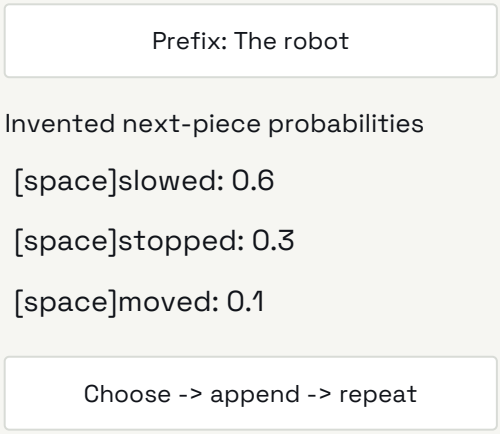
For Sam's summary, imagine the prefix "The robot". Our invented next-piece distribution is [slowed]: 0.6, [stopped]: 0.3, and [moved]: 0.1. These probabilities sum to one over this deliberately tiny toy vocabulary. They are not probabilities that the corresponding events happened.

A decoding rule selects a token. Choosing the largest probability and sampling are different rules; sampling can vary continuations from the same model. [T14] Choosing " slowed" expands the prefix for the next step.

A fluent continuation can still misstate the note or invent an event. Learned language patterns can reproduce false beliefs, as tested in TruthfulQA; its historical benchmark results do not rank today's systems. [T12]

Check Sam's summary against the note and recompute 12/15. Token probabilities describe text; they do not replace source inspection or arithmetic. Different sampled answers still need evaluation.

THE MECHANISM



D05. A toy three-piece vocabulary, summing to one. Selection is not independent verification of an event.

SOURCES T15 · T14 · T12 / LINKED IN THE TEXT COMPANION

CHECK YOUR UNDERSTANDING

Does a 0.6 next-token probability mean a 60% chance the robot actually slowed?

Answer: No. It is a probability over the next token in the specified toy model.

09 / INSIDE AI

Reasoning behavior needs evaluation

Distinguish useful multistep performance from a trustworthy explanation.

Language models can perform useful multistep tasks. Research has found that prompting with intermediate steps can improve performance on tested arithmetic and other tasks. This is evidence about behavior under specified conditions, not a guarantee for every problem or proof of a human-like inner process. [T07]

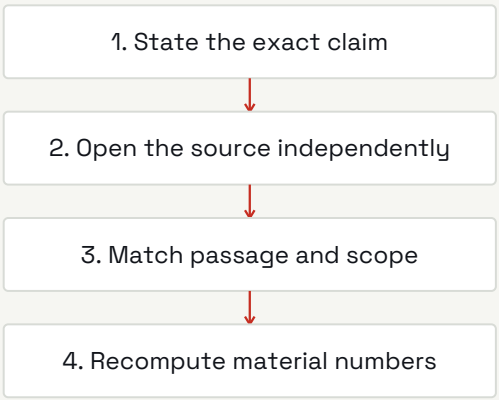
Suppose Sam receives: "12 successful trials out of 15 is 90%, so reliability is excellent." The prose is decisive, but the arithmetic is wrong. Independently, 12/15 = 0.8 = 80%. Three trials were unsuccessful. Even a corrected percentage describes those trials, not safety in all future conditions.

Inspect visible steps without assuming they report every influence. Experiments found explanations omitting input biases affecting predictions. This establishes a failure possibility, not that every explanation is false. [T08]

Tool use can make checking easier: a calculator can compute a fraction and a search tool can find a source. Check the actual tool input, result and interpretation. A tool being available is different from evidence that it ran correctly on this task.

Specify the question, inputs and acceptable result before testing. Try an unfamiliar comparable problem. Score correctness and evidence, not length.

THE MECHANISM



Supported / unresolved / contradicted

Record limits and next evidence.

D06. Original audit method. A source can exist without supporting the generated conclusion.

SOURCES T07 · T08 / LINKED IN THE TEXT COMPANION

CHECK YOUR UNDERSTANDING

Which supports the 80% result more directly: a long explanation or independently dividing 12 by 15?

Answer: The independent calculation; neither alone establishes general robot safety.

10 / INSIDE AI

Know the boundary of the evidence

Identify context, freshness, generalization, evaluation and privacy limits.

Sam's note supports a narrow statement about a small fictional test. To judge a new use, ask what changed. A robot tested in a quiet corridor may face different lighting, people or surfaces elsewhere. Evidence from one setting does not automatically transfer to another. [T10]

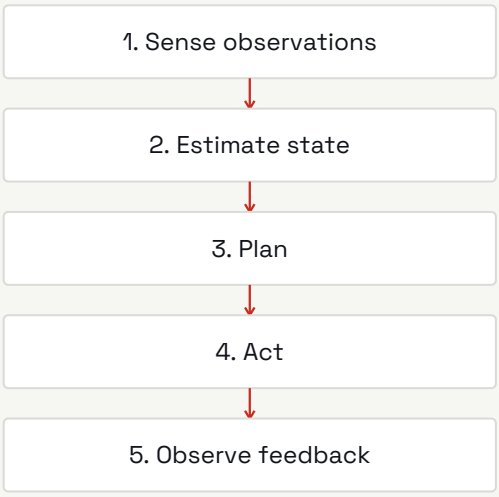
Ask what information was actually available. If the assistant was not given a later incident report and did not retrieve it, the answer has no demonstrated basis in that report. A context limit also bounds what can be supplied in a single model input; a system's memory and retrieval features are separate mechanisms. [T15, T09]

Distinguish a benchmark from the intended job. A score measured on one question set is evidence for that set and evaluation procedure. It does not certify all tasks, populations or real-world conditions. Keep successes, failures and the total tested count. [T11]

Before sharing work documents, check authorization and actual data handling. Remove unnecessary sensitive material. Privacy depends on the whole system and its use. [T11]

State Sam's boundary: "The note reports 12 stops in 15 trials; it does not establish readiness for deployment around people."

THE MECHANISM



Feedback informs the next cycle.

Check delay, uncertainty and limits.

D07. A conceptual loop. Language plans need grounding in physical capabilities; this diagram is not a safety architecture.

SOURCES T10 · T15 · T09 · T11 / LINKED IN THE TEXT COMPANION

CHECK YOUR UNDERSTANDING

What evidence is missing for an unseen crowded environment?

Answer: Relevant tests and system evidence for those conditions; the corridor note alone is insufficient.

11 / INSIDE AI

A citation is a claim to check

Verify a reference and explain retrieval limits.

An AI answer may contain confidently presented false or unsupported content, often called hallucination or confabulation. A reference formatted like a real paper is still a claim needing inspection. NIST identifies this risk and recommends reviewing sources and citations. [T11]

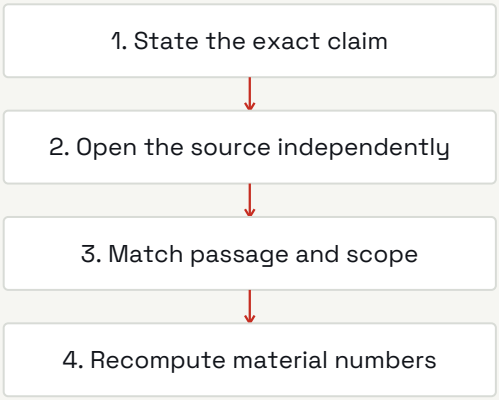
Here is an intentionally invented teaching reference: "Example Author, 2026. Perfect Robot Safety from Fluent Conversation. Journal of Imaginary Machines, 0(0), 1-2." This is fabricated for the exercise. It is not evidence, and no real publication or author is being accused of anything.

Search independently for the publication. Verify authors and date, open it and locate the supporting passage. Check this exact claim and comparable conditions; a working link is insufficient.

Retrieval-augmented generation supplies retrieved material to a generation system. It can improve factuality on evaluated tasks, but retrieval may miss the relevant source, select poor material or be misinterpreted in the answer. Lewis and colleagues' results show improvements, not perfect factuality. [T09]

Finding a sentence does not establish a conclusion. Neither this fabricated reference nor a language-model paper substitutes for physical-system evidence. Mark unverifiable claims unverified.

THE MECHANISM



Supported / unresolved / contradicted

Record limits and next evidence.

D06. Original audit method. A source can exist without supporting the generated conclusion.

SOURCES T11 · T09 / LINKED IN THE TEXT COMPANION

CHECK YOUR UNDERSTANDING

A real linked paper discusses robot planning. Does that establish Sam's robot is safe?

Answer: No. Verify the exact claim, system, test conditions and limitations.

12 / INSIDE AI

From chat to action in the world

Name the embodied feedback loop and its extra evidence needs.

A chatbot can describe an action. A physical system must connect decisions to an environment. A useful engineering abstraction is: sense -> estimate state -> plan -> act -> observe feedback. Cameras or other sensors supply observations; estimation interprets them; planning and control connect goals to actions. [T10]

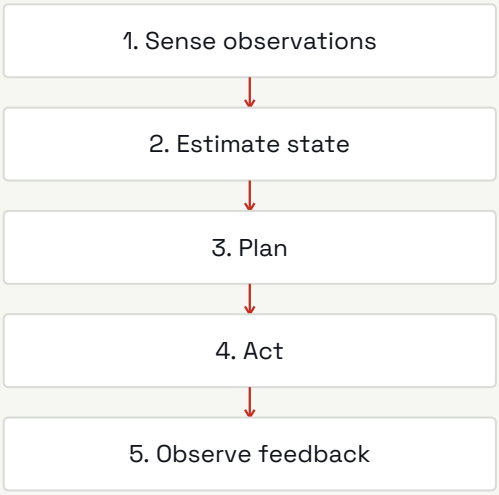
A sensor may detect something near the robot's path. Estimation interprets what and where it might be. Planning may call for slowing; control must make the hardware respond. Feedback checks the result while the world changes.

Latency, missed observations, uncertain estimates and limited physical capabilities can break a plausible plan. Research connecting language models to robots explicitly grounds language choices in available skills and their feasibility. Its particular experiments do not certify other robots. [T10]

A multimodal system can accept more than text, such as images. That alone does not supply a tested actuator, feedback loop or physical safety case. Evaluate the complete system and its environment, including failure handling. Our diagram is conceptual, not instructions for operating a robot.

Sam can use an assistant to organize questions. Claiming safe operation needs engineering evidence beyond this guide; fluent conversation alone does not demonstrate physical competence.

THE MECHANISM



Feedback informs the next cycle.

Check delay, uncertainty and limits.

D07. A conceptual loop. Language plans need grounding in physical capabilities; this diagram is not a safety architecture.

SOURCES T10 / LINKED IN THE TEXT COMPANION

CHECK YOUR UNDERSTANDING

What must happen after issuing a command?

Answer: Observe the result through feedback and respond appropriately; issuing text alone is insufficient.

13 / INSIDE AI

Audit an AI output in one page

Apply a reusable audit and answer three self-checks.

Use these five prompts before relying on an output. This is an original educational checklist; it is not a certification or a complete safety review.

01 Task: What exact claim am I using? Separate the supplied data, a calculation and a broader conclusion.

02 Mechanism: Was this generated from context, retrieved evidence or a tool result? What was actually available?

03 Evidence: Open the primary source, find the supporting passage and recompute material quantities independently.

04 Stakes and limits: What changes in the intended setting? Who is affected if it is wrong? What expert or system evidence is needed?

05 Decision: Accept a bounded claim, revise it, reject it or leave it unresolved. Record why and what would change the decision.

Worked audit: Sam can accept that the supplied fictional note reports 12/15 successful stops, after checking the text and 80% calculation. Sam cannot infer safe deployment from that count. Missing evidence includes the test design, conditions, failure severity and relevant system validation.

01 Self-check 1: Is token ID 42 a vector? Answer: no; it identifies an entry that can select a vector.

02 Self-check 2: Does a plausible explanation verify 12/15 = 90%? Answer: no; independent arithmetic gives 80%.

03 Self-check 3: Does a chatbot's safe-sounding plan establish physical competence? Answer: no; sensing, control, feedback and relevant testing remain necessary.

Try the audit: keep one verified statement and one uncertainty.

CHECK YOUR UNDERSTANDING

What is Sam's bounded conclusion?

Answer: The note reports 80% successful stops in those trials; broader safety is unresolved.

14 / INSIDE AI

Keep the distinctions, follow the sources

Use the glossary, sources and contact without an availability promise.

Token: a text piece in a tokenizer's vocabulary. ID: its identifying index.

Embedding: a numerical vector selected or produced to represent an item. Hidden state: an intermediate representation in a computation.

Parameter: a stored model value, such as a weight. Activation: a value computed for particular inputs.

Training: adjustment of parameters against an objective. Inference: computation using a model on inputs.

Attention: a weighted combination of value representations. Logit: an unnormalized output score; it is not yet a probability.

Autoregressive: generating successive tokens conditioned on preceding context. Retrieval: fetching external material.

Hallucination: generated false or unsupported content. Embodied system: a system acting through a physical body with environmental feedback.

Start with Vaswani et al., Attention Is All You Need [T01], and Jurafsky and Martin, Speech and Language Processing, 19 August 2026 draft [T02]. For the limits, see Jain and Wallace on attention [T06], Turpin et al. on explanations [T08], Lewis et al. on retrieval [T09], and Ahn et al. on robotics [T10]. The companion contains every linked reference and source scope.

I'm Aldiyar, director of Enterprise AI Adoption at Swiss Innovators Club, with experience in Switzerland's industrial technology sector and a Master's degree in electrical engineering and IT from ETH Zurich. I'm a member of Mensa Switzerland and created OnePrompt to make AI, robotics and emerging technologies easier to understand without sacrificing technical depth.

Contact: aldiyar@swissinovatorsclub.ch. Instagram and LinkedIn: semedyarov. Version 1.0, 28 October 2026. Reply with Understand, Apply or Build to describe your interest. This guide does not announce course availability or a certificate.

SOURCES T01 · T02 · T04 · T06 · T08 · T09 · T10 · T11 · T15 / LINKED IN THE TEXT COMPANION

CHECK YOUR UNDERSTANDING

Where can you check a source or report a correction?

Answer: Use the linked references and author contact in the companion.