

Ourbit Vendor-Neutral AI - Day 2: How Generative AI and Language Models Work

Day 2 - How Generative AI and Language Models Work

Programme: Vendor-Neutral AI - 15-Day Certification Programme

Purpose: Understand enough of language-model generation to interpret outputs intelligently without depending on one vendor's terminology or interface.

Learning outcomes

- Explain token-by-token generation in plain language.
- Distinguish a likely continuation from a verified fact.
- Explain why supplied context changes an answer.
- Recognize why repeated runs can vary.
- Apply verification when a claim matters.

1. From text to tokens

Language models process text as units commonly called tokens. A token may correspond to a word, part of a word, punctuation or another text fragment. The important lesson is not the exact tokenizer: the model is working with encoded units and learned statistical relationships rather than reading sentences with human understanding.

2. Generation is sequential

Given the available context, a language model estimates plausible next tokens. One token is selected, added to the context, and the process continues. Repeated many times, this produces paragraphs, code or other structured text. The resulting response can appear coherent because the model has learned rich patterns in language and data.

But “plausible next token” is not the same operation as “look up and prove a fact.” A model may generate a statement that fits the linguistic context even when the statement is inaccurate or unsupported.

3. Why context matters

The instruction, conversation, supplied documents and other available context shape what continuations are plausible. If critical information is missing, ambiguous or contradictory, the response may reflect that weakness. Better context can improve relevance, but adding more text indiscriminately can also introduce noise.

4. Why outputs can vary

Generation is probabilistic. More than one continuation may be plausible, and systems can be configured to favor more predictable or more varied outputs. Therefore, identical requests do not necessarily produce identical answers. Variation can be useful for ideation but becomes a design consideration when consistency matters.

5. Generation versus evidence

For creative work, plausibility may be enough. For factual or consequential work, important claims need evidence appropriate to the task. That may mean checking an authoritative source, comparing source documents, using a controlled retrieval system, running a calculation, or asking a qualified human. Verification is not distrust of AI; it is good information practice.

Worked example - the polished statistic

Imagine a model produces: “The market grew by 37% last year,” with no source. The number's precision can make it feel authoritative. Yet the generation mechanism alone provides no proof that 37% is correct, that “market” has the intended definition, or that “last year” matches the required period. A responsible workflow treats the claim as unverified until appropriate evidence supports it.

Guided practice

Ask any available language model or use a hypothetical response if no tool is available to explain a topic containing a factual claim. Mark each sentence as: general explanation, externally verifiable claim, inference, or recommendation. Identify which statements need evidence before being used in a report.

Knowledge check

- What role do tokens play in language-model generation?
- Why can a plausible continuation still be factually wrong?
- How can context improve or weaken an answer?
- Why might two responses to the same prompt differ?
- When should an external verification step be added?

Applied evidence task

You receive an AI-written briefing containing three precise statistics and two named claims but no citations. Create a verification plan. State which claims you would check first, what kind of source would be appropriate, how you would record uncertainty, and what you would do if reliable evidence could not be found.

Day 2 takeaway

Language models are powerful generators of plausible sequences. Their usefulness increases when users understand the difference between generation and evidence, provide relevant context, expect some variation and verify important claims.

Assessment boundary

This day contributes only to vendor-neutral-ai-15-day. Learners are assessed on transferable understanding, not knowledge of a particular model provider.

Learning comes before credentials. Completing this reading alone does not issue a certificate. Programme assessment and practical requirements apply.