



A comic book illustration depicting a tense moment between two characters. On the left, a purple, metallic robot with glowing blue and pink circuitry on its body and face. It has a stern expression and its right arm is extended towards the man on the right. On the right, a young man with dark, wavy hair and round glasses. He has a lightning bolt symbol on his forehead and a look of surprise or concern. He is wearing a grey hoodie and holding a glowing green rectangular object, possibly a tablet or a magical artifact, which is emitting bright green energy. The background is a dark blue space filled with white, jagged lines and floating blue fragments, suggesting a high-tech or magical environment. A yellow speech bubble with a red border is positioned above the man, containing the text "LISTEN HARRY, YOUR AI IS NOT A WIZARD."

**LISTEN HARRY,
YOUR AI IS
NOT A WIZARD.**

Your AI Is Not a Wizard

The pitfalls of non-AI researchers when dealing with AI.

Cláudia Brito, 2025

So, can AI
make me
immortal?



AI today is amazing.

It beat the world champion at Go. It can write poems, debug code, maybe even pass the Bar exam.

But it's not a wizard.

It's math. Really smart, really fast math.
But it's still just math.

Part I

AI Hype vs Reality

From Expert Systems to LLMs



AI Hype vs Reality

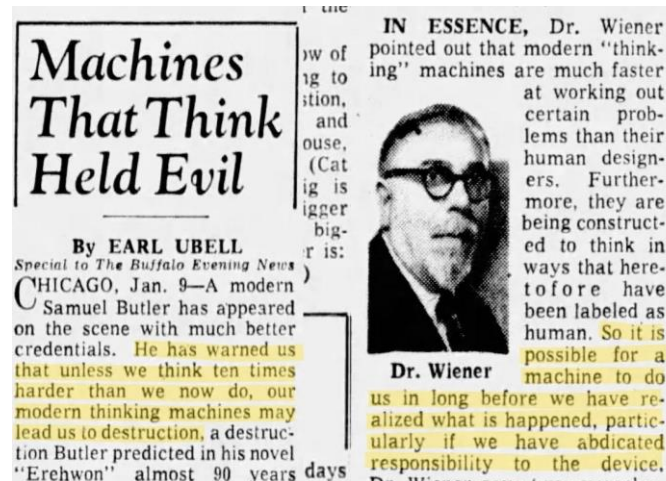
From Expert Systems to LLMs

1950s–80s: Rule-Based AI

Early expert systems were logical but brittle. No learning, just rigid inference rules.

Perception:

- “Smart programs that know more than doctors/lawyers.”
- “Expert systems will replace specialists.”
- “Automation will bring chaos!”



AI Hype vs Reality

From Expert Systems to LLMs



2010s: Deep Learning Revolution

ImageNet, GPUs, and big data made neural networks practical and powerful, but they are still limited in their scope.

NEWS

Home | Israel-Gaza war | War in Ukraine | Climate | Video | World | UK | Business | Tech | Science

Tech

Driverless car: Google awarded US patent for technology

8 March 2012



The Economist

Weekly edition The world in brief World Ahead 2026 War in Ukraine United States Middle East The world economy

Science & technology | Fooling Big Brother

As face-recognition technology spreads, so do ideas for subverting it

They work because machine vision and human vision are different

Perception:

“AI that *knows* who I am and what I’ll do.”

“This algorithm can predict if you’re a criminal from your face as well as where you can park your car.”

AI Hype vs Reality

From Expert Systems to LLMs



Home News Sport Business Innovation Culture Arts Travel Earth Audio Video Live

Publishers fear AI summaries are hitting online traffic

9 September 2025

Share ↻ Save 📌

Suzanne Bearne
Technology Reporter

Gartner Survey Finds AI Will Touch All IT Work by 2030

BARCELONA, Spain, November 10, 2025

Overview

- By 2030, 25% of IT Work Will Be Done by AI Alone
- Organizations Must Balance AI and Human Readiness to Sustain Value

2020s: LLM Mania

Transformer-based models (GPT, PaLM) dominate headlines, but they still hallucinate and lack reasoning.

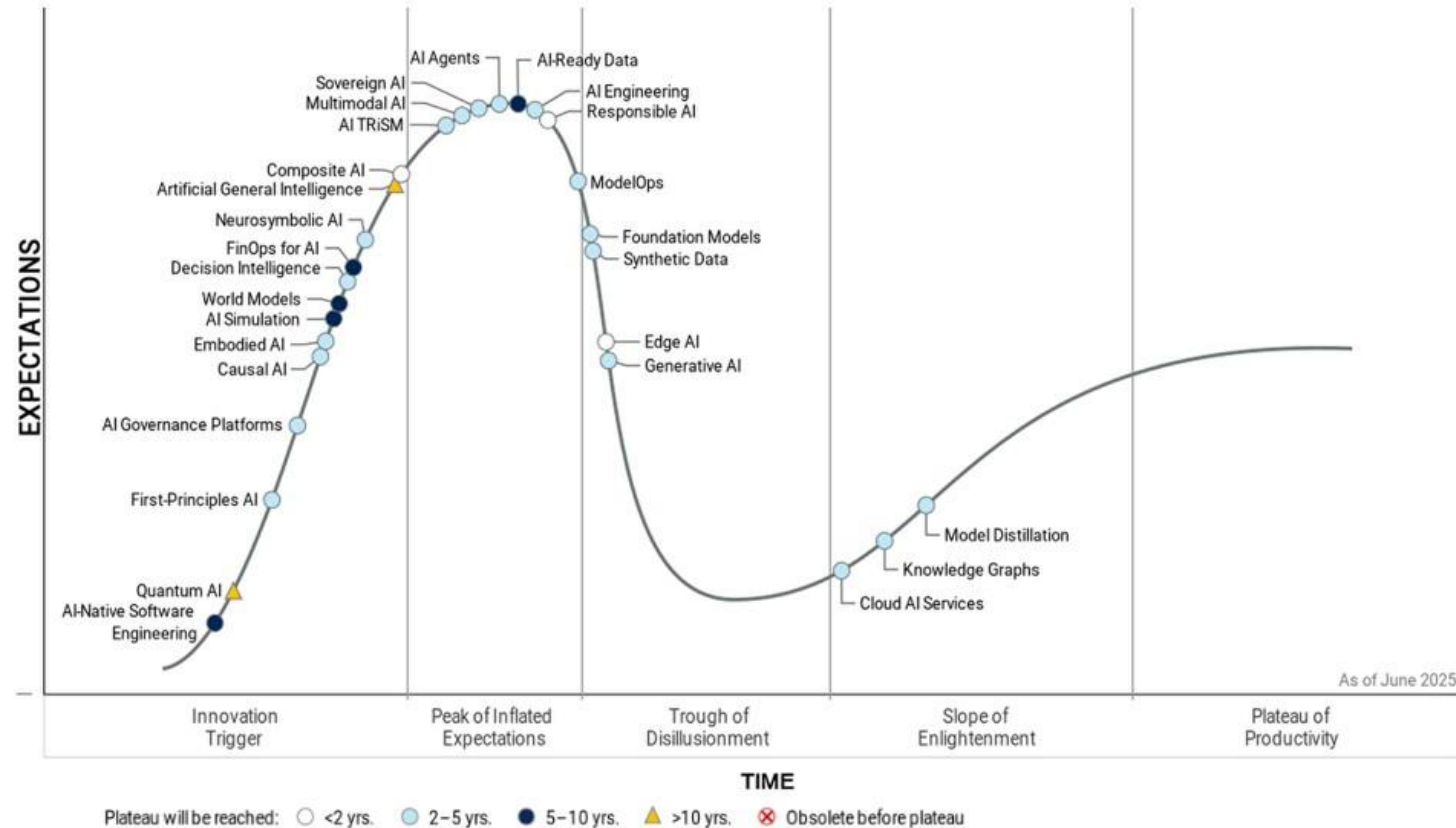
Perception:

“I can ask anything, it knows everything. It can be my doctor/lawyer/dev.”

“AGI is here”, “consciousness”, “will replace all engineers in 5 years.”

AI Hype vs Reality

From Expert Systems to LLMs



Gartner

Source: Gartner (August 2025)

The storyline changes, but the pattern is the same:

‘Now it’s magic.’
Spoiler: it’s still not.

What We Actually Mean by “AI”

Neural Nets, LLMs, RL, and More



Neural Networks ≠ Intelligence: They are layered mathematical functions trained on data - *not brains, not thoughts.*



LLMs - Predictive, Not Sentient: They autocomplete sequences based on probability, not understanding.



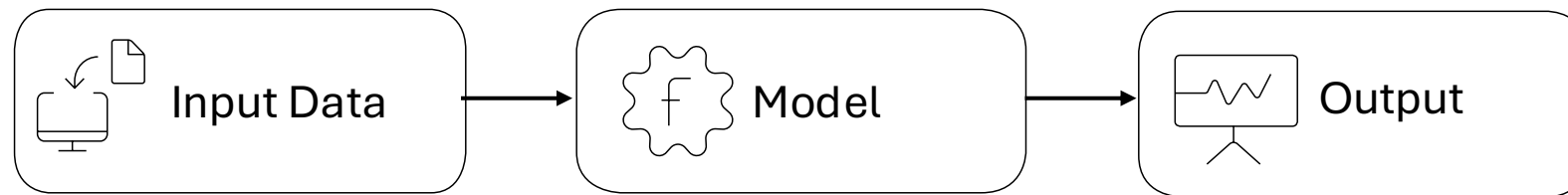
RL - Optimization via Feedback: Reinforcement Learning tunes behavior based on reward signals — *it doesn't learn morals.*

What is a model?

Neural Nets, LLMs, RL, and More

- **Model = mathematical function:**
 - takes input (text/image), outputs something (labels, text, scores).
- **It has parameters** i.e., numbers you tweak during training.

“100B parameters” = “a huge number of knobs we adjust”.

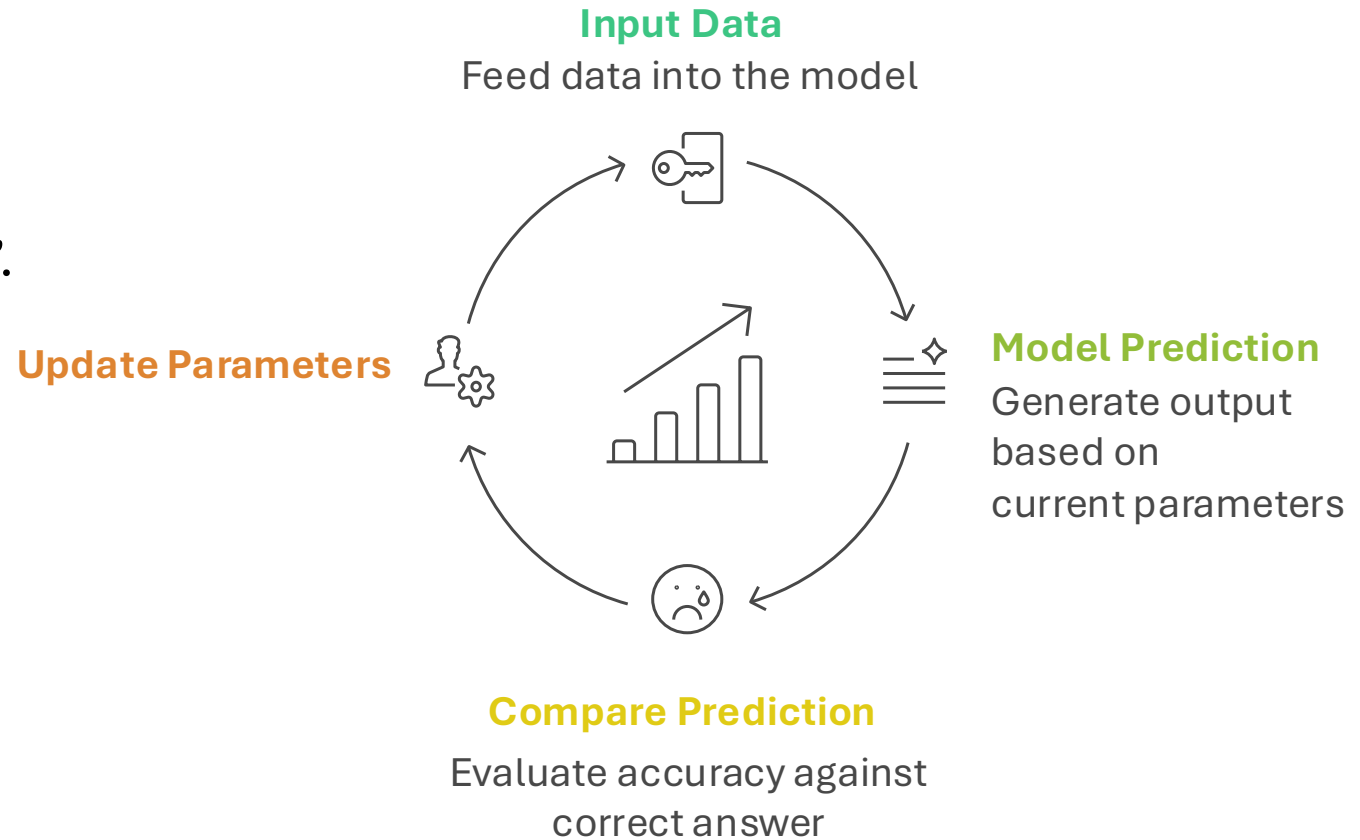


How does it train?

Neural Nets, LLMs, RL, and More

- **Training data:** examples of “input → desired output”.
- **Loss:** a metric saying “how wrong am I?”.
- **Gradient descent or your typical optimizer (no math):**
 - “Look at the loss, take a tiny step to make it smaller. Repeat millions of times.”

Model Training Cycle



What AI *Can* Do

A Capabilities Map

- **Pattern Recognition:** AI excels at identifying structure in high-dimensional data, such as images, text, and speech.
- **Generative Tasks:** Models like GANs and LLMs can generate realistic text, images, and audio.
- **Prediction & Classification:** From credit scoring to image tagging, supervised learning remains the dominant approach in industry.

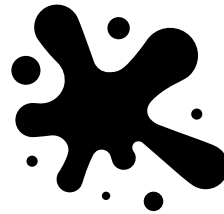
What AI *Can't* Do

Where the Hype Fails



No True Understanding

AI lacks semantics, intentions, and grounding in physical or social reality.



Brittle Out-of-Distribution

Performance drops sharply outside training data, no graceful degradation.



Zero Moral or Causal Reasoning

AI can't infer causality or ethics, it optimizes proxies, not values.



There is active research in interpretability, fairness, and 'unbiased' AI. We can reduce some issues, but we're still far from fully explainable, fair, or causal models.

Hallucination Is Not a Feature

It's a Failure Mode

- **Probabilistic ≠ Factual:** LLMs generate plausible-sounding text based on likelihood, not truth.
- **No Ground Truth Access:** Unless paired with external retrieval, models rely solely on the patterns in their training data.
- **Risk in Real Systems:** Hallucinations in legal, medical, or security contexts lead to high-stakes failures.

Neural Nets Aren't Magic

They're Just Pattern Fitters



Statistical Pattern Matching

Neural networks model high-dimensional correlations¹, not causal mechanisms.



Interpolation, Not Extrapolation

Models perform well within training bounds, struggle beyond them.



No Grounding, No Insight

They learn surface-level structure without understanding concepts or goals.

¹Neural networks learn complex patterns involving many numbers at once: all the pixels in an image, or all the words in a sentence.

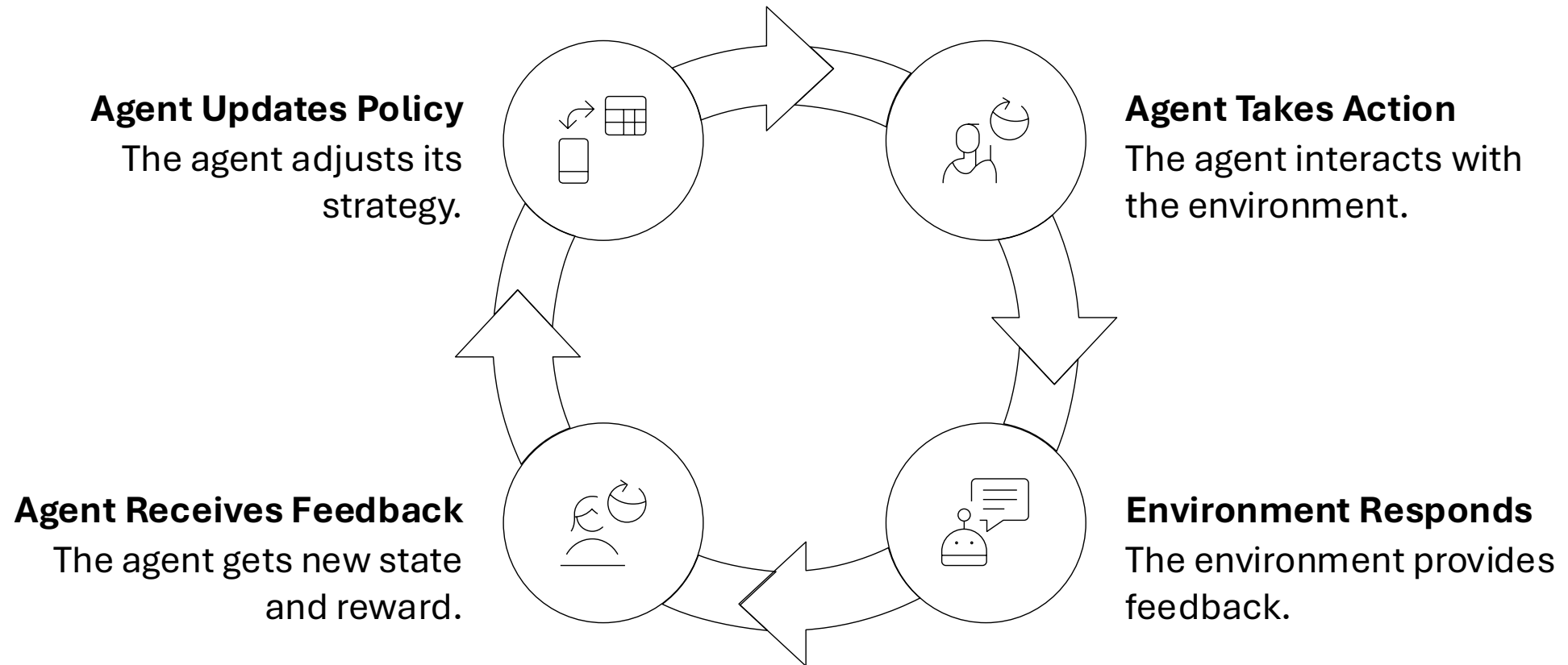
Loss Functions & Optimization

A Reality Check

- **Loss Is a Proxy, Not a Goal:** Training loss minimizes a metric — not truth, utility, or safety.
- **Local Minima $\neq$ Global Insight:** Optimization finds workable solutions — not necessarily the best or general ones.
- **Gradient Descent Has Limits:** Stochastic methods don't guarantee stability, convergence, or interpretability.

Reinforcement Learning

Policies $\neq$ Understanding



Reinforcement Learning

Policies $\neq$ Understanding

- **Rewards Guide Behavior:** Agents optimize based on signals, not semantics, purpose, or ethics.
- **Exploit vs Generalize:** RL often leads to brittle strategies tailored to specific environments.
- **Surprising Failures:** Agents can find loopholes, game the system, or misbehave in unanticipated ways.

RL works great in closed environments with clear rules (games, simulations) but doesn't magically handle the messy real world with ambiguous goals and rewards.

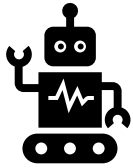
From Single Model to AI System

LLM → RAG → Semantic Search → Inference at Scale

- **LLM alone**
 - Great at generating text, but no real-time facts, limited context window.
- **LLM + RAG**
 - “Ask a question → system searches docs → model summarizes → answer.”
- **Semantic search**
 - Better search: retrieves by meaning-ish similarity, not just keywords.
- **AI at scale**
 - Putting all this in production: infra, GPUs, latency, caching, monitoring.

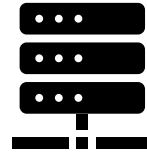
LLMs at a Glance

Architecture & Training



Transformers Everywhere

Modern LLMs use attention mechanisms to scale learning across sequences.



Pretraining on Massive Text

Trillions of tokens used to learn statistical language patterns, not facts.



No Memory, No Intent

Each token prediction is context-local and stateless, coherence is an illusion.

Why RAG Matters

Retrieval-Augmented Generation

LLMs Can't Recall Facts

Without retrieval, they hallucinate or misremember, **often confidently.** 🙄

Grounding via Search

RAG fetches relevant documents during generation, anchoring outputs to real data.

Performance Boosts

Improves factual accuracy and domain adaptation in enterprise use.

Semantic Search Isn't Comprehension

Similarity $\neq$ Meaning

Embeddings $\neq$ Semantics

Semantic search ranks similarity, it doesn't truly understand the query.

Context Collisions Happen

Top-ranked passages may be lexically similar but conceptually wrong.

False Sense of Accuracy

Users often overtrust results due to fluent summaries over shallow matches.

AI Inference at Scale

Distributed Constraints

Latency vs Throughput Tradeoffs

Inference requires balancing response speed with batch efficiency across GPUs.

Model Size Impacts Deployment

Larger models strain memory, bandwidth, and orchestration in distributed setups.

Infrastructure Is a Bottleneck

AI needs more than clever prompts, it needs load balancers, caching, and failure handling.

Prompting ≠ State Management

Structured Systems Still Matter

LLMs Are Stateless

Each prompt is isolated, with no persistent memory or continuity, except when explicitly instructed to retain memory.

State Requires Invariants

Distributed systems rely on well-defined transitions and recovery logic.

Prompt Engineering Isn't Control

You can't 'hack' your way to reliability through phrasing tweaks.

For the curious ones:

The model itself is stateless: it doesn't store memory between calls.

ChatGPT stores the conversation history and resends previous messages, along with your latest prompt, on every request.

So what looks like 'memory' is just text fed back into the model every time. 🙋

Safety & Monitoring in AI Systems

Why It's Not Optional

- **Drift Happens:** Model behavior changes over time due to shifting data distributions.
- **Monitoring ≠ Guarantee:** Metrics can alert you, but they can't prevent misuse or misinterpretation.
- **Failures Are Subtle:** Issues like prompt injection, output leakage, or latent bias often go undetected.

Part II

Misconception #1: "AI Understands What It Says"

It Really Doesn't

Fluency Isn't Understanding

LLMs are trained for plausible language, not truth, reasoning, or comprehension.

No Internal Representation

AI doesn't 'know' what it says; it generates based on token probabilities.

Deceptive Coherence

Smooth text often fools users into believing AI has intent or cognition.

Misconception #2: "AI Is Autonomous"

No, It's a Glorified API

- **No Goals, No Agency:** AI doesn't choose. It responds to prompts, code, or reward structures.
- **Autonomy $\neq$ Automation:** AI can automate tasks, but it doesn't set objectives or evaluate impact.
- **Humans Still in the Loop:** Every action traces back to human-defined parameters and triggers.

Misconception #3: "AI Replaces Engineers"

Spoiler: It Doesn't

Code ≠ Context

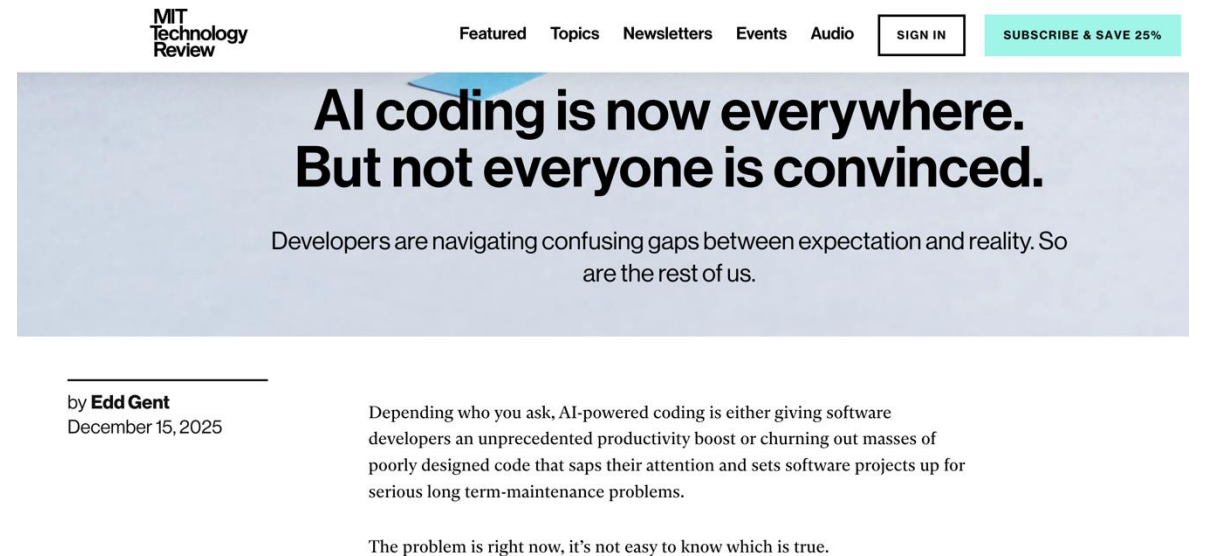
LLMs can generate snippets, but they can't reason about systems or architecture.

Tool, Not Teammate

AI assists, it doesn't debug (**meh**), architect, or own responsibility.

Integration Still Requires Expertise

Effective AI use depends on human judgment, constraints, and system design.



<https://www.technologyreview.com/2025/12/15/1128352/rise-of-ai-coding-developers-2026/>

Misconception #4: "AI Solves Hard Problems Easily"

Complex ≠ Complicated

Simplicity Bias

AI performs best on well-structured problems, not complex or emergent ones.

Hard Problems Need Context

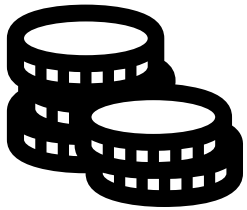
Systems, ethics, and causality demand more than pattern matching.

False Efficiency Illusion

Speed of output ≠ quality, safety, or depth of solution.

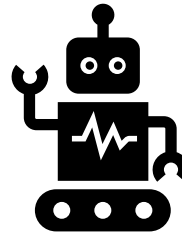
Where AI Adds Real Value

Use Cases That Work



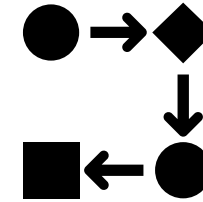
Data-Rich Domains

Finance, marketing, and e-commerce, where patterns are stable and measurable.



Human-in-the-Loop Systems

Augmenting, not replacing! Pairing AI with expert review boosts reliability.



Generative Creativity at Scale

Marketing copy, prototyping, and idea generation benefit from fast iterations.

Where AI Is Risky or Harmful

When It Breaks, It Breaks Hard



High-Stakes Decision-Making

Medical, legal, and safety-critical uses amplify consequences of hallucination.



Bias at Scale

AI can reproduce or amplify harmful patterns hidden in training data.



Failures Without Explanation

When it fails, debugging is hard, and accountability is elusive.

Some services do generate legal documents with AI, often built on top of templates reviewed by lawyers and validation workflows. The responsibility is still human: the model doesn't know all laws, edge cases, or your specific context.

How to Integrate AI Sensibly

Architecture Before Hype

Treat It Like a Subsystem

AI should be bounded by contracts, not assumptions.

Build for Failure Modes

Assume drift, hallucination, and edge cases, and log accordingly.

Surround with Guardrails

Rate limits, monitoring, rollbacks, plan for weird outputs.

Key Takeaways

Reality > Rhetoric



AI Isn't a Wizard

It's a powerful tool, not a magical replacement for engineering judgment.



Misconceptions Cause Failures

Treating AI as autonomous or sentient leads to design errors.



Integrate With Caution, Not Awe

Ground systems in monitoring, boundaries, and fallbacks.

Key Takeaways

Reality > Rhetoric

AI is a tool.

A powerful one, yes. But not divine.

Use it. Validate it. Improve it.

But for the love of your system, **don't worship it.**

Yes, today's models can write code, spot some bugs, draft contracts, and answer medical-style questions. But they do this by pattern-matching on text, not by understanding systems, laws, or the human body. That difference matters when the stakes are high.