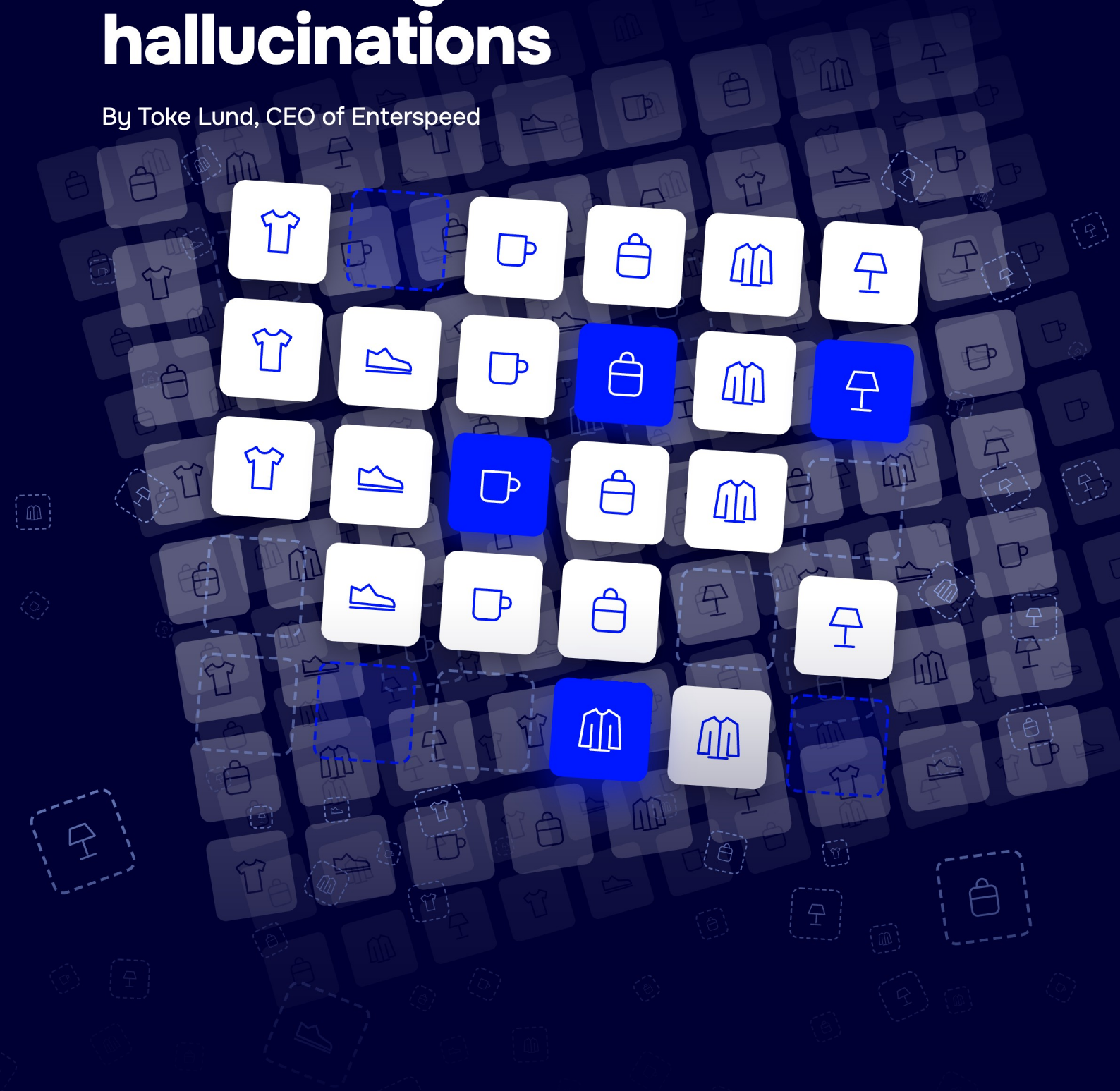
Enterspeed

WHITE PAPER FOR RETAILERS

10 things about controlling AI and its hallucinations

By Toke Lund, CEO of Enterspeed



Published April 2026. Updated November 2026.

Contents

Intro 3

01 Not all context is equal 4

02 Human-on-the-loop: oversight that scales 5

03 Your training data is only half the battle 6

04 Data governance prevents hallucinations 7

05 Standardisation is your friend (even when it's boring) 9

06 Not all hallucinations are created equal 10

07 Monitoring and feedback loops are non-negotiable 12

08 Explainability is a control mechanism 13

09 The temperature setting everyone forgets 14

10 Trust, but verify, at every level of the architecture 15

What this means in practice 16

So, what should you do next? 17

About Enterspeed 18



INTRO

“How do we trust AI?”

The question I hear most often these days.

Every retailer I've met that uses large language models has seen hallucinated output: a product description that invents a specification, campaign text that contradicts brand guidelines, or customer-facing copy with the wrong price. Errors like these show up in every enterprise that scales AI, which makes hallucinations an architecture question.

AI hallucinations come from how language models work: they predict the most statistically likely next word rather than the most factually correct one. What you can control is what the model predicts from. Controlling hallucinations therefore means designing the environment AI works in, above all the context it receives.

Weak context is also why most AI projects stall after the pilot. The Context Factory is how we fix that. It turns the data you already have into precise context, in three steps: **data in, context out, AI at scale**. This paper covers ten things that work when you deploy AI at enterprise scale, written for the people who have to make it work.

Toke Lund
CEO of Enterspeed

This edition adds Enterspeed's own benchmarks and results from XL-BYG, the Danish builders' merchant that checks its product content with Speedtrain.

01 / Not all context is equal

The most effective way to reduce hallucinations is to give AI the right data at the right time. Enterprises often either flood models with irrelevant information (creating noise) or starve them of necessary context (forcing the model to guess).

For a retailer, this plays out every day. Your AI is generating product descriptions, but it's pulling from a data model that hasn't been updated since last season. Or it has access to everything (deprecated SKUs, old supplier copy and conflicting specifications included) and doesn't know which source to trust.

The fix is better-prepared data. Your AI needs structured, validated and context-rich data, prepared in advance and served to match each question. For retail, that means product data, pricing, stock and content read from one governed truth over the systems you already run. Your ERP and PIM stay authoritative. The same goes for product feeds: Google Shopping and each marketplace need their own shape, built from the same prepared data as the product page, not from a separate export.

ENTERSPEED BENCHMARKS

COST AND SPEED

The same enrichment task, run on raw product data and then with the product taxonomy prepared up front.

300,000 → 15,000

tokens used

5× cheaper

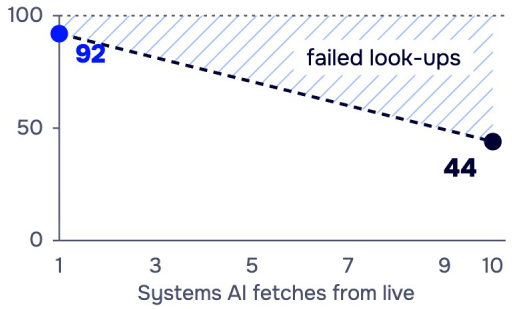
3× faster

Same quality

A token is the unit of text a model reads and is billed for.

RELIABILITY

How often an AI agent's look-ups work, out of 100, as it fetches live from more systems.



Enterspeed's own measurement. Method available on request.

Preparing the data up front pays off twice. The AI reads less, so each answer costs less and arrives faster. And it reads from one place, so fewer things can fail along the way. When the AI answers from one prepared version of your authoritative data, it has far less to guess.

02 / Human-on-the-loop: oversight that scales

You can't have a human approve every AI output when you're processing thousands of product descriptions, campaign texts or customer interactions daily. But abandoning oversight entirely isn't the answer either.

The right model is "human-on-the-loop": intelligent flagging systems that identify which outputs need human attention and let the rest flow through. Think of it like quality control in manufacturing. You don't inspect every unit. You build processes that catch anomalies and flag edge cases.

STEP 1
Every output
Product content and answers, at scale

STEP 2
Automatic checks
Confidence and risk flags from the data

STEP 3
Only the flagged
reach a person. The rest flows through.

In practice, this means confidence scoring: the system knows when a model is uncertain, outside its training domain or handling high-stakes outputs, like compliance-sensitive product claims.

You can only act as a human-on-the-loop if your data carries the metadata to flag risks automatically. Without proper data architecture, you're back to manual review or no review at all. That takes tagging, structuring and enriching your data so the system knows what's sensitive, what's verified and what needs a human eye.

IN PRACTICE: XL-BYG

XL-BYG, the Danish builders' merchant, runs its product content this way. Speedtrain from Enterspeed generates and checks each product text. Only the texts that fail a check go to a product specialist, and in XL-BYG's setup 99 percent pass automatically.

30 min → 2 min per product

"Our product specialists validate rather than write. That uses their expertise better."

Kim Portz Knudsen, Head of Webshop, XL-BYG

The 99 percent reflects XL-BYG's workflow and data quality. A typical first run approves around 60 percent automatically.

03 / Your training data is only half the battle

Training data quality matters, but production AI systems also have to stay current. This is particularly acute in retail, where product ranges, pricing, promotions and compliance requirements change constantly.

Your beautifully trained model from six months ago doesn't know about this quarter's product launches, pricing changes or new compliance requirements.

Retrieval-augmented generation (RAG) solves this. Instead of relying purely on what the model "remembers" from training, you dynamically inject current, relevant information at the moment of generation. The AI then reflects your current business rather than a snapshot from months ago.

What your AI knows, with and without RAG



RAG is only as good as your retrieval system. If you're pulling irrelevant documents or outdated information, you've traded one problem for another. The retrieval mechanism needs to be intelligent, fast and deeply integrated with your business logic. That depends on data architecture. When a specification is corrected in the PIM, every text built on it should follow, in every language, instead of someone fixing six translations by hand.

04 / Data governance prevents hallucinations

Imagine your AI assistant:

serving up product information for a line you've discontinued,

generating content from a supplier spec sheet superseded six months ago,

or using pricing data from a market you no longer serve.

These are compliance problems, and they are also organisational hallucinations, where the AI's output conflicts with your actual business reality.

Proper data governance creates boundaries that prevent AI from accessing or surfacing information it shouldn't. This means implementing real-time policy controls that understand data sensitivity, user permissions and regulatory requirements. An AI agent should see only what the person or system asking is entitled to see.

When these guardrails are built into your data, you prevent compliance violations and a whole class of hallucinations where AI makes up responses because it lacks proper context about what it can and can't reference. The data can also say what it doesn't know, so "unknown" becomes an answer the AI can give instead of a guess.

The same question, and the care instructions are missing from the product data:

WITHOUT GOVERNANCE

Is this jacket machine washable?

Yes, at 40°C.

A likely-sounding guess. If the jacket is dry-clean only, it gets ruined and returned.

WITH GOVERNANCE

Is this jacket machine washable?

I don't have care instructions for this jacket. Please check the label.

The data marks care instructions as unknown, so the AI says so instead of guessing.

The enterprises doing this well have moved beyond static access controls to dynamic, context-aware data policies.

Poor governance creates legal risk and AI that contradicts your own business.

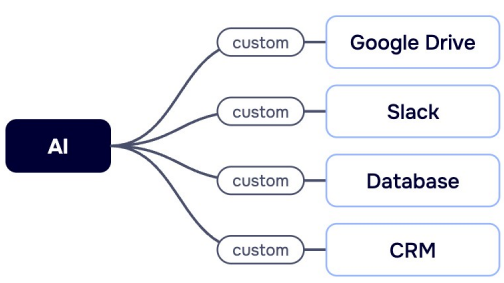
05 / Standardisation is your friend (even when it's boring)

The dirty secret of enterprise AI is that a lot of teams are rebuilding the same things: custom integrations for every data source, bespoke connectors for every tool and one-off implementations that are impossible to scale. This fragmentation wastes engineering time and creates hallucination vectors, because every integration is a potential point of failure.

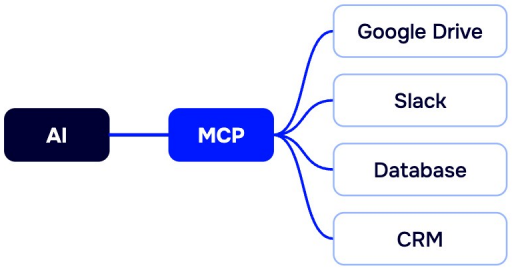
Without standardisation, you're building legacy debt with every new AI integration.

Each custom connector becomes technical debt that someone has to maintain, update and eventually replace. Worse, these one-off solutions make it nearly impossible to implement consistent validation and governance across your AI systems.

Open standards like the Model Context Protocol (MCP), which defines one way for AI to connect to data and tools, are changing this dynamic. Instead of writing custom code to connect your AI to Google Drive, then Slack, then your database, then your CRM, you build against a single protocol. Think USB-C for AI: one standard that works everywhere.



ONE CUSTOM CONNECTOR PER SYSTEM



ONE STANDARD FOR ALL

This standardisation matters for hallucination control because it allows you to implement consistent validation, monitoring and governance across all your data sources.

When you have uniform interfaces, you can build uniform safety measures. Plus, the time you save on integration work can be redirected to what actually matters: prompt engineering, testing and validation.

06 / Not all hallucinations are created equal

There's a crucial distinction between creative hallucinations and factual hallucinations. In brainstorming, campaign ideation and product naming, some speculation is actually useful.

The problem arises when you need factual accuracy, but your system is configured for creativity. This is why temperature settings, system prompts and model selection matter so much. A customer service bot should be configured conservatively, but a marketing ideation tool can run hotter.

Smart enterprises don't deploy one-size-fits-all AI. They build task-specific configurations with controls matched to the risk, down to single fields: a compliance claim gets required wording and forbidden terms, while a campaign tagline gets rules for tone and brand.

Match the controls to the risk: creative tasks can run loose, compliance content cannot.

A content brainstorm tool and a product compliance tool should not have the same guardrails.

USE CASE	TYPICAL CONTROLS	TOLERANCE	RISK
Creative / Exploratory Ideation, brainstorming, marketing concepts	Loose prompts, high temperature	**High**	**Low**
Informational / Operational Internal assistants, support responses	RAG, confidence scoring, source validation	**Medium**	**Medium**
Critical / Regulated Legal, financial, medical, compliance	Human approval, audit logs, strict governance	**Near zero**	**High**

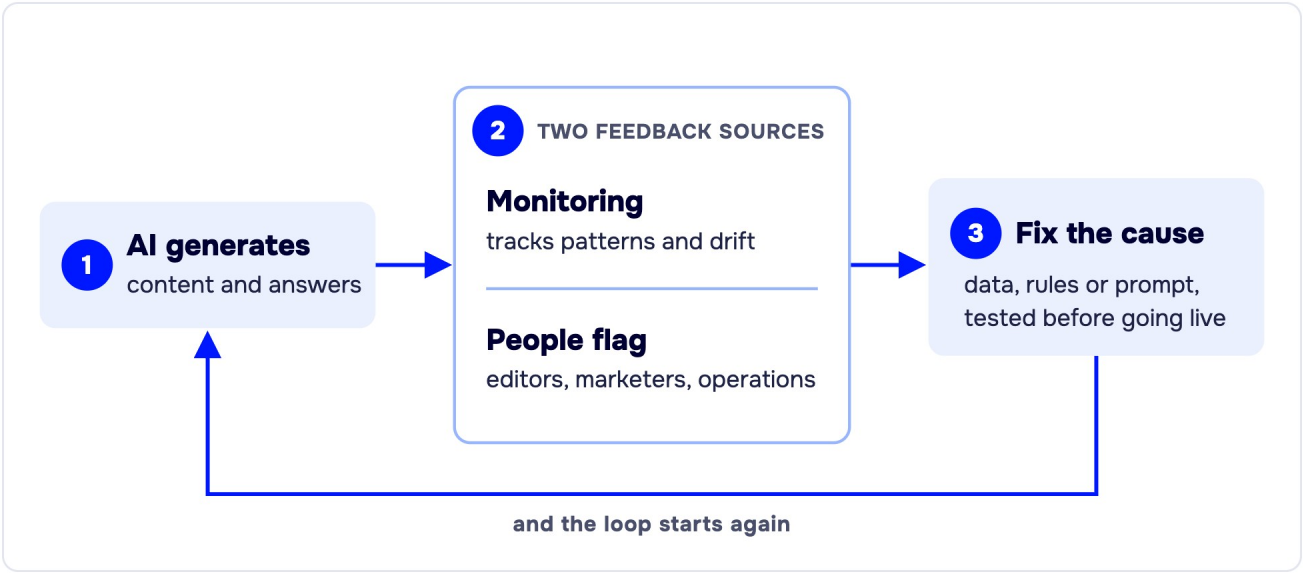
For retail, product compliance content, legal text and pricing sit firmly in the "critical" category, and campaign ideation and draft copy in the "creative" one. Treating them the same way is where things go wrong.

Your product feed is critical too. Google Merchant Center rejects a product when a required attribute is missing, and may do so when the price doesn't match the product page, often without anyone noticing until sales drop. Get it right and the feed pays back: **after enriching its product content, XL-BYG saw 12% more clicks from Google Shopping.**

07 / Monitoring and feedback loops are non-negotiable

Monitoring AI outputs means more than logging requests and responses. It means tracking hallucination patterns, measuring output quality drift over time and identifying which types of queries consistently produce unreliable results.

The best systems implement real-time telemetry that detects when model behaviour starts diverging from expected patterns, before it reaches customers or gets published. This requires building feedback mechanisms where editors, marketers and operations teams can flag incorrect outputs, which then feed into improvement cycles. Version the rules your AI works by, too, so a change can be tested against the previous version before it becomes the default.



You wouldn't deploy a new ecommerce feature and never monitor its performance. AI systems need the same discipline – arguably more, because AI behaviour can drift in ways that traditional software doesn't.

If you're not monitoring your AI outputs systematically, you're flying blind.

08 / Explainability is a control mechanism

When you can trace why an AI produced a particular output (which data sources it weighted, which context it prioritised), you can spot hallucinations faster.

This becomes particularly powerful when combined with domain expertise. A product manager who can see which parts of the data the model weighted heavily can quickly identify when it's over-indexing on irrelevant information.

In retail, explainability matters at every level: compliance teams need audit trails, editors need to understand why a description was generated a certain way, and leadership needs confidence that the system is operating as intended. In practice, that means recording for every output where each value came from and which rules it passed, so guesses become something you can count.

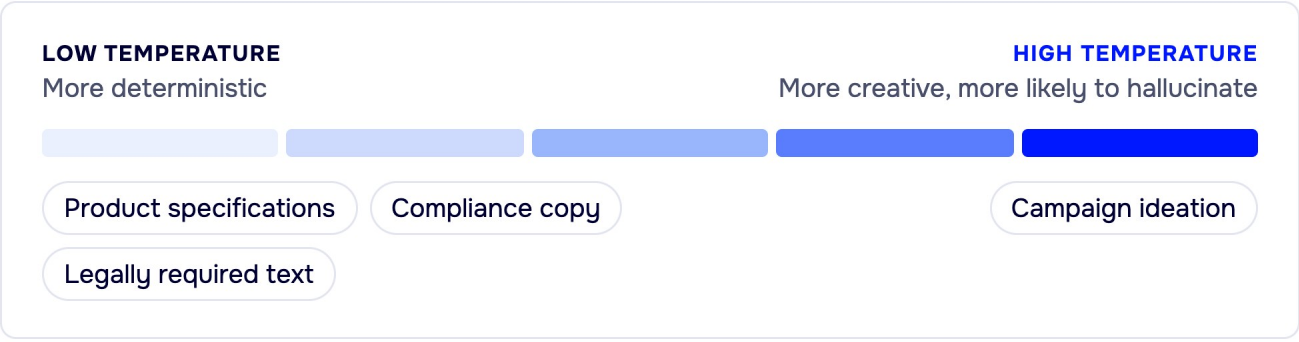
Black box versus glass box AI

DIMENSION	BLACK BOX AI	GLASS BOX AI
Visibility	None	**High**
Data sources	Unknown	**Explicit and traceable**
Decision logic	Opaque	**Interpretable**
Error detection	Reactive	**Proactive**
Human oversight	Difficult	**Built-in**
Regulatory readiness	Low	**High**
Trust level	Assumed	**Verifiable**

When you understand why AI made a decision, you can catch errors before they compound.

09 / The temperature setting everyone forgets

Model temperature controls the creativity-accuracy trade-off. Higher temperatures make models more creative, and more likely to hallucinate. Lower temperatures make them more deterministic, which is exactly what you want when generating product specifications, compliance copy or legally required text.



The mistake most teams make is using default settings across all use cases. Your product content tool should run conservatively: a low temperature where the model allows it, and strict rules where it does not. Your campaign ideation tool can run much higher. These aren't just technical parameters – they're decisions about how much creative latitude your AI is allowed to take with your brand.

The key is matching temperature settings to your specific use case and then testing rigorously. Find the sweet spot between useful creativity and reliable accuracy for each application, and document it as part of your AI governance framework.

Temperature isn't a technical detail. It's a business decision about acceptable risk.

10 / Trust, but verify, at every level of the architecture

The fundamental principle underlying all of this: no matter how good your controls are, you need verification mechanisms baked into your architecture. At every layer.

The retailers moving fastest with AI right now aren't the ones who've found a way to trust AI blindly. They're the ones who've built layered systems that catch and correct errors before they reach a product page, a campaign or a customer. They architect for failure, because they know failure will happen.

Trust is built not by assuming correctness, but by verifying it at every level. On the next page, you'll see what that looks like when you translate it from principle into practice.

The enterprises succeeding with AI aren't the ones who've eliminated hallucinations. They're the ones who catch them before they cause damage.

THE TEN THINGS AT A GLANCE

01 Not all context is equal

02 Human-on-the-loop: oversight that scales

03 Your training data is only half the battle

04 Data governance prevents hallucinations

05 Standardisation is your friend (even when it's boring)

06 Not all hallucinations are created equal

07 Monitoring and feedback loops are non-negotiable

08 Explainability is a control mechanism

09 The temperature setting everyone forgets

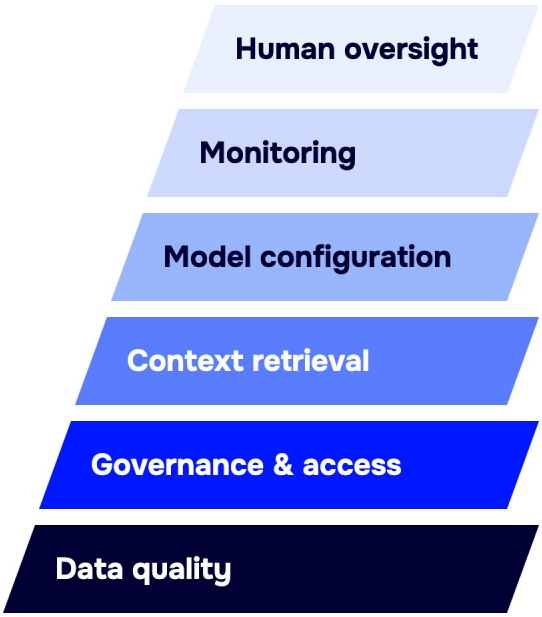
10 Trust, but verify, at every level of the architecture

What this means in practice

- Automated validation against authoritative data sources for factual claims
- Confidence scoring that surfaces uncertain outputs for review
- Human approval workflows for high-stakes, customer-facing content
- Audit trails so you can trace exactly how an output was generated
- Staged rollouts where new AI capabilities are tested at limited scale before full deployment

Defence-in-depth for AI hallucination control

Read from the bottom: each layer catches what the one below lets through.



LAYER	PURPOSE	RISK ADDRESSED
Human oversight	Final accountability	High-stakes errors
Monitoring	Detect drift & patterns	Silent degradation
Model configuration	Control creativity	Over-speculation
Context retrieval	Ground answers	Knowledge gaps
Governance & access	Enforce boundaries	Unauthorised data use
Data quality	Trust the inputs	Garbage-in hallucinations

Hallucination mitigation cannot rely on a single control.

Effective systems apply layered defences across data quality, governance, context injection, model configuration, monitoring and human oversight.

Each layer reduces risk, and together they create resilient, auditable AI architectures.

So, what should you do next?

The good news: you don't need to reinvent the wheel. The tools, standards and architecture patterns exist. What's needed is the discipline to treat data quality and governance as strategic priorities, not IT hygiene.

The companies getting this right are discovering that the same infrastructure that reduces hallucinations also makes their AI faster, more accurate and more valuable. Better data architecture reduces risk and lets AI do much more for your business.

Trustworthy AI isn't about having perfect models. It's about having the right systems in place for when they're imperfect. And they will always be imperfect.

START WITH THREE QUESTIONS

01 **Where is your AI currently generating outputs without a reliable, real-time data source to ground it?**

02 **Which of your AI use cases carry the highest business risk if they hallucinate, and do they have the right guardrails?**

03 **Do you have monitoring in place that would tell you if output quality started to degrade?**

THE CONTEXT FACTORY **Data in. Context out. AI at scale.**

Data in
Every source, from products, prices and stock to supplier files and PDFs, kept fresh in real time.

Context out
Connected, gaps filled, errors fixed and adapted to each market. Your team stays on the loop and only sees the exceptions.

AI at scale
The same precise context for your webshop, feeds, marketplaces and AI agents, and back to your PIM.

If you're not sure where to start, we'd be happy to walk you through how other retailers are approaching this, including how XL-BYG moved from writing product content to validating it.

Get in touch

info@enterspeed.com
+45 21 55 18 25

About Enterspeed

Enterspeed is the Context Factory. We turn the data you already have into precise context, so AI makes it out of the pilot and into production.

We take in data from the systems you already run, refine it with AI under your rules, and serve it as one precise context to your webshop, your feeds and your AI agents. Your team handles the exceptions.

TRUSTED BY ENTERPRISES INCLUDING

XL BYG · Carlsberg · Danish Crown · hummel · Flügger · whiteaway · Novicell

ISO/IEC 27001 ISMS

Enterspeed's Information Security Management System is certified to ISO/IEC 27001 by Insight Assurance, and Enterspeed is GDPR compliant.

Data in. Context out. AI at scale.

- **Data in.** We read the systems you already run, from PIM and ERP to supplier files and PDFs, without an integration project or a migration.
- **AI at scale.** We serve the same precise context to your webshop, feeds, marketplaces and AI agents, and your team can ask what's wrong with the data in Claude or ChatGPT.
- **Context out.** We write and check product content with Speedtrain, against your own rules, and return it to the PIM you already use.
- **Control.** AI agents get read-only access, so they see only what their caller is entitled to see, and your systems stay closed.

Enterspeed

Data in.
Context out.
AI at scale.

Get in touch with us anytime.

info@enterspeed.com

+45 21 55 18 25

Enterspeed A/S, P.O. Pedersens Vej 2
DK-8200 Aarhus N, Denmark