

MODULE 1

# Context – From Data Chaos to Clarity

Production architecture deep-dives and live reviews

**Daniel Töws**  
codecentric

**Alexander K.**  
Horsch

**Gauthier Salavert**  
AskUnali

Agents without  
proper context  
are just more  
expensive  
hallucinations

# Agenda

- 18:00 - 18:05 - Welcome
- 18:05 - 18:15 - Recap & Opening
- 18:15 – 18:45 - Talk 1: [Daniel Töws](#), codecentric AG
  - LLMs vs. Messy Data – Lessons from the Trenches
- 18:45 – 19:15 - Talk 2: [Alexander K](#), Horsch
  - plot-RAG (pRAG): Visualizing & Optimizing RAG Performance – Smarter RAG Evaluation
- 19:15 - 19:45 - Startup Challenge: [Gauthier Salavert](#), AskUnali
  - Hybrid RAG
- 19:45 – 21:30 - Closing & Networking

# Why Context Matters (From Last Meetup)

- 42% said context is the most overlooked layer.
- 41% said data chaos is their biggest blocker.
- Today: we dig into *why*.

“Context failures look like AI failures.”

# What's Killing Production AI (Your Response)

Primary Cluster	%
Fragmented data sources	33%
Evaluating context quality	24%
Data quality vs delivery	22%
Debugging agent reasoning	21%

But the free-text questions go much deeper.  
Tonight, we use those to drive the discussion.

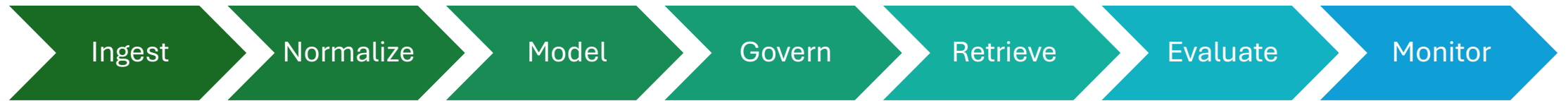
# What You *Actually* Care About

Six dominant themes from free text:

1. Evaluation & Trust
2. Data Quality, Validation & Lineage
3. Retrieval Stability & Context Bloat
4. Context Infrastructure (contracts, versioning, governance)
5. Security, Access Control, Agent Identity
6. Maintainability & Autonomy

Far beyond RAG demos — this is real production architecture.

# The Context Lifecycle

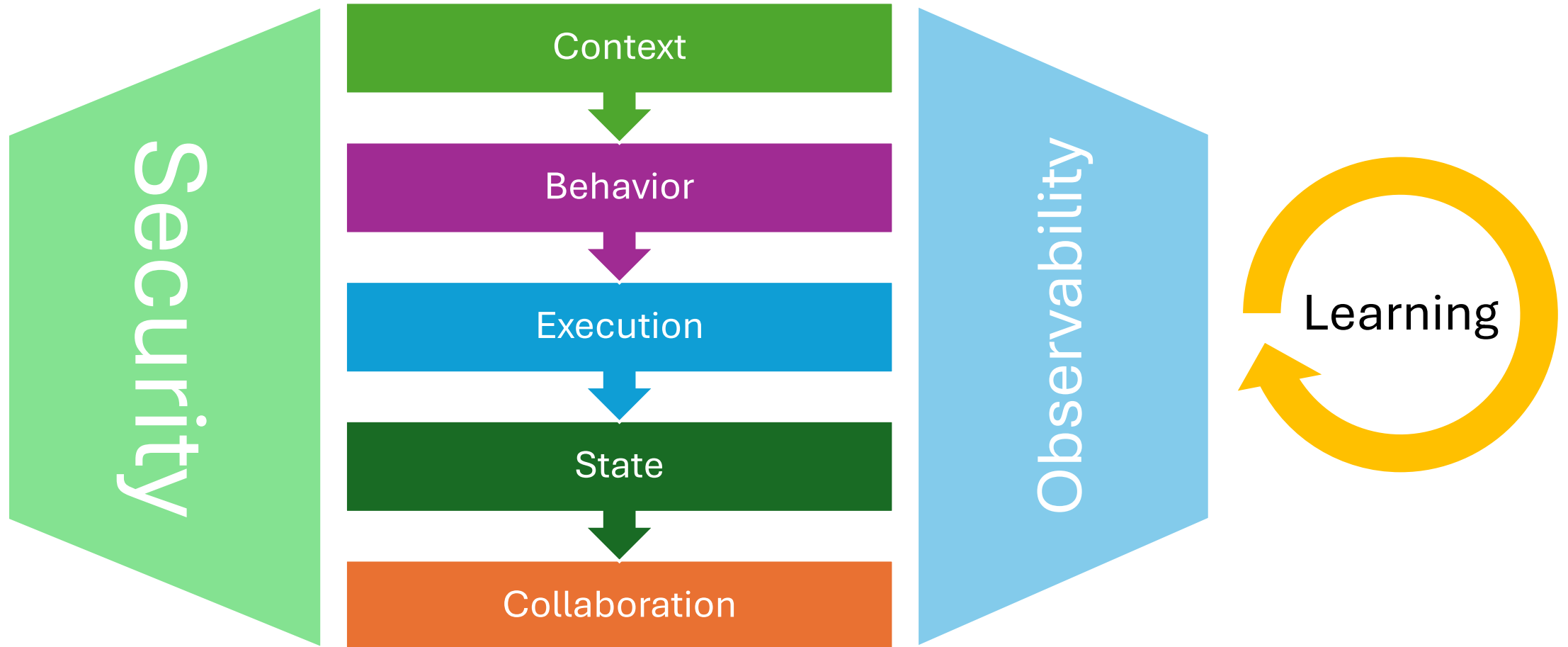


Every failure you reported sits somewhere here.

# CAA: The Blueprint for Production-Grade “Agents”

- Context — coherent, typed, versioned inputs
- Behavior — explicit, inspectable plans
- Execution — deterministic workflows
- State — persistent memory
- Collaboration — approvals + overrides
- Cross-cutting: Observability & Security  
Lifecycle: Learning & Adaptation

# These Are Not Model Problems — They're CAA Layer Failures





# Challenge Cluster → CAA Layer Mapping

	Cluster A	Cluster B	Cluster C	Cluster D
Challenge	<b>Fragmented or inconsistent data sources</b>	<b>Data quality vs. delivery speed</b>	<b>Evaluating context quality (RAG)</b>	<b>Debugging agent reasoning</b>
CAA Layer	<b>Context layer</b>	<b>Context + State + Observability</b>	<b>Context + Observability + Learning</b>	<b>Behavior + Observability + Collaboration</b>
Pain	Missing structure, canonicalization, contracts	Missing validation, tests, lineage, monitors.	Lack metrics, evaluation loops, continuous improvement	Hidden plans, no inspectability, no approval gates

# Mapping Your Themes → CAA Layers

- **Evaluation & Trust** → Observability, Behavior, Learning
- **Data Quality & Lineage** → Context, State
- **Retrieval Stability** → Context, Execution, Observability
- **Context Infrastructure** → Context, State, Security
- **Access/Identity/Security** → Security, Context, Collaboration
- **Maintainability & Autonomy** → Execution, Collaboration, Learning

# Cluster – Talks Mapping

	Cluster A	Cluster B	Cluster C	Cluster D
Theme	<b>Fragmented or inconsistent data sources</b>	<b>Data quality vs. delivery speed</b>	<b>Evaluating context quality (RAG)</b>	<b>Debugging agent reasoning</b>
Talk	Daniel (Talk 1)	Daniel (Talk 1)	Alexander (Talk 2)	Startup Challenge

# Diagnostic Exercise (2 minutes)

Think of the last time your AI system failed in production (or in a prototype). One specific failure, not a category.  
Write one sentence about what went wrong.

# Locate the true root cause using CAA

- **Context:** broken inputs, missing structure, inconsistent meaning
- **Behavior:** hidden plans, unclear reasoning, bad decomposition
- **Execution:** tools misfiring, workflows non-deterministic
- **State:** missing memory, stale memory, no persistence
- **Collaboration:** missing approval gates, no oversight
- **Observability:** no trace, can't debug, invisible failure
- **Security:** PII leaks, unsafe data flows, isolation gaps
- **Learning & Adaptation:** system doesn't improve, drifts

# Identify the mechanism of failure

- Write down the mechanism that caused that layer to fail.  
Examples: missing typing, no versioning, hidden reasoning, nondeterministic tool calls, stale memory, no human approval, no monitoring, no eval loop.

# Lock it in

- Keep that root-cause layer in mind.  
You'll see it reflected in every talk tonight.

# Mentimeter





# Talk 1 — Daniel Töws

# Talk 2 — Alexander Kress (pRAG)

# Startup Challenge (Hybrid RAG)

Gauthier - AskUnali

# Diagnosing From Sparse Symptoms

- Curated medical dataset with enriched metadata
- Clinics query API with keyword-like symptom descriptions
- Sometimes additional metadata (age, sex)
- System must return:
  - 1 highly likely cause
  - 1–2 plausible / surprising causes
- Input often extremely sparse (input quality degrades in multi-turn)
- Challenge: **weighting & ranking with incomplete context**

# Dimension A — Signal Weighting Under Sparse Input

- How should you weight symptom embeddings vs metadata vs disease priors?
- Should you use an LLM rewrite/expansion step to “enrich” sparse queries?
- Should you dynamically choose retrieval strategy based on input richness?

# Dimension B — Ranking Strategy for Medical Suggestions

- How to combine relevance, likelihood, risk, and “clinical usefulness”?
- How to enforce surprise vs expected causes (1 likely, 1 unlikely-but-relevant)?
- Should the system use a scoring model? Heuristic? LLM judge? Hybrid?

# Dimension C — Context Filtering & Governance

- Should the system use strict hard filters (sex/age)?
- Should there be explainability constraints on why a cause is returned?
- How to prevent over-fitting to metadata?
- How to avoid degraded suggestions after multiple turns?

# Case Study — When Context Fails

Oct 1, 2025, 15:02 GMT+2

Hallo Herr Müller,

Um Ihnen so schnell wie möglich weiterzuhelfen, erhalten Sie eine Antwort von [REDACTED], unserem digitalen Assistenten:

Vielen Dank für Ihre Nachricht bezüglich der fehlenden Rechnungen für Ihr Abo [REDACTED].

Ich verstehe, dass Sie seit August keine Rechnungen mehr erhalten haben. Als Unternehmen sollten Sie normalerweise monatlich eine Sammelrechnung für alle abonnierten [REDACTED] erhalten.

Alle Rechnungen sind im [REDACTED] [Business Portal](#) hinterlegt, wo Sie diese einsehen und herunterladen können. Falls Sie noch keinen Zugang zum Business Portal haben oder Probleme beim Einloggen auftreten, könnte dies der Grund sein, warum Sie die Rechnungen nicht erhalten.

Bitte beachten Sie auch, dass jedem Unternehmen eine individuelle IBAN zugeordnet ist und Sie abhängig vom Vertragspartner ([REDACTED] GmbH) gegebenenfalls zwei Rechnungen pro Monat erhalten könnten.

Um das Problem zu lösen, empfehle ich Ihnen folgende Schritte:

1. Überprüfen Sie Ihren Zugang zum [REDACTED] Business Portal
2. Kontrollieren Sie Ihren E-Mail-Spam-Ordner für eventuelle Rechnungen
3. Stellen Sie sicher, dass Ihre Kontaktdaten aktuell sind

Können Sie mir bestätigen, ob Sie Zugang zum [REDACTED] Business Portal haben und ob sich Ihre E-Mail-Adresse oder andere Kontaktdaten kürzlich geändert haben?

Wir danken Ihnen für Ihre Geduld und Ihr Verständnis. Antworten Sie auf diese E-Mail, wenn Sie eine weitere Fragen haben.

Ihr [REDACTED] Team

Um Ihnen so schnell wie möglich weiterzuhelfen, erhalten Sie eine Antwort von [REDACTED], unserem digitalen Assistenten:

I understand you don't have access to the Business Portal. This explains why you haven't been receiving your monthly invoices since August for subscription [REDACTED].

For business customers, monthly invoices are typically available through the [REDACTED] Business Portal. You can find an overview of your monthly invoices by selecting "Dokumente" in the main menu on the left side of the portal. From there, you can search for specific invoice numbers and download them by clicking the "Download" button under "Aktion".

Since you currently don't have portal access, this is likely preventing you from accessing your invoices. Each company has a specific IBAN assigned for settling outstanding invoices, which you can find on your invoices. Please note that the IBAN differs per company, so if you're a customer with multiple subsidiaries, you'll have a separate IBAN for each.

To resolve this issue, you'll need to get your Business Portal access set up so you can:

1. View and download your monthly invoices
2. Access the correct IBAN information for payments
3. Ensure your contact details are current

Without portal access, you won't be able to retrieve the missing invoices from August onwards or access the payment details needed for your subscription.

Wir danken Ihnen für Ihre Geduld und Ihr Verständnis. Antworten Sie auf diese E-Mail, wenn Sie eine weitere Fragen haben.

Ihr [REDACTED] Team

Hinweis: Bitte beachten Sie, dass die Aussagen von [REDACTED] nicht bindend sind.



# How This Maps to the 9 Modules

- Context — tonight
- Behavior — next session
- Execution
- State
- Collaboration
- Observability
- Security
- Learning & Adaptation
- Scaling & ROI

**We're building the shared  
architecture for real-world AI  
systems.**

<https://github.com/Munich-AI-Nexus/production-readiness-blueprint/>