



Trustworthy AI for Food Safety: An Architecture for Verifiable, Accountable Decision Support

29.09.2026, KIDA KON 2026 Berlin

Stefan Dehm¹, Alexandra Fetsch², Martin Wiedmann³, Iurii Savvateev², YeonJin Jung³,
Caroline Renea Yates³, Monika Ehling-Schulz⁴, Sophia Johler², Matthias Filter¹

¹ Innovation Centre Food Chain Modelling and Artificial Intelligence, Federal Institute for Risk Assessment, Berlin, Germany, ² Competence Centre for Food Safety, Chair of Food Microbiology, Veterinary Faculty, LMU Munich, Germany, ³ Department of Food Science, College of Agriculture and Life Sciences, Cornell University, Ithaca, USA, ⁴ Center of Pathobiology, Institute of Microbiology, University of Veterinary Medicine, Vienna, Austria

In the food industry, simple questions can be hard to answer

Complex environment – complex requirements

EU, FDA, national laws, export markets, retailer standards, best practices, research results, competing stakeholder interests – all applying to one product

Regulation, procedure and evidence are all provided in natural language

Language models are a good fit for a decision support system

But: language models are inherently unreliable

→Further grounding is required

Can we extend shelf life by two days if the cold chain holds at 6°C?

We have a positive finding. Do we need to initiate a recall?

How often should we test our work surfaces for *Listeria spp.*?

Empowering LLMs for the task at hand

Integrated Components for trustworthiness.

Knowledge base

Every answer is grounded and carries a citation.
Unsupported answers are discarded.

Predictive models

Recommending and executing a FSKX-simulation grounds the answer further.

Risk negotiation

Transparently balancing competing stakeholder interests and valuation of trade-offs.

User context

User, role and facility profiles shape and customize the answer.

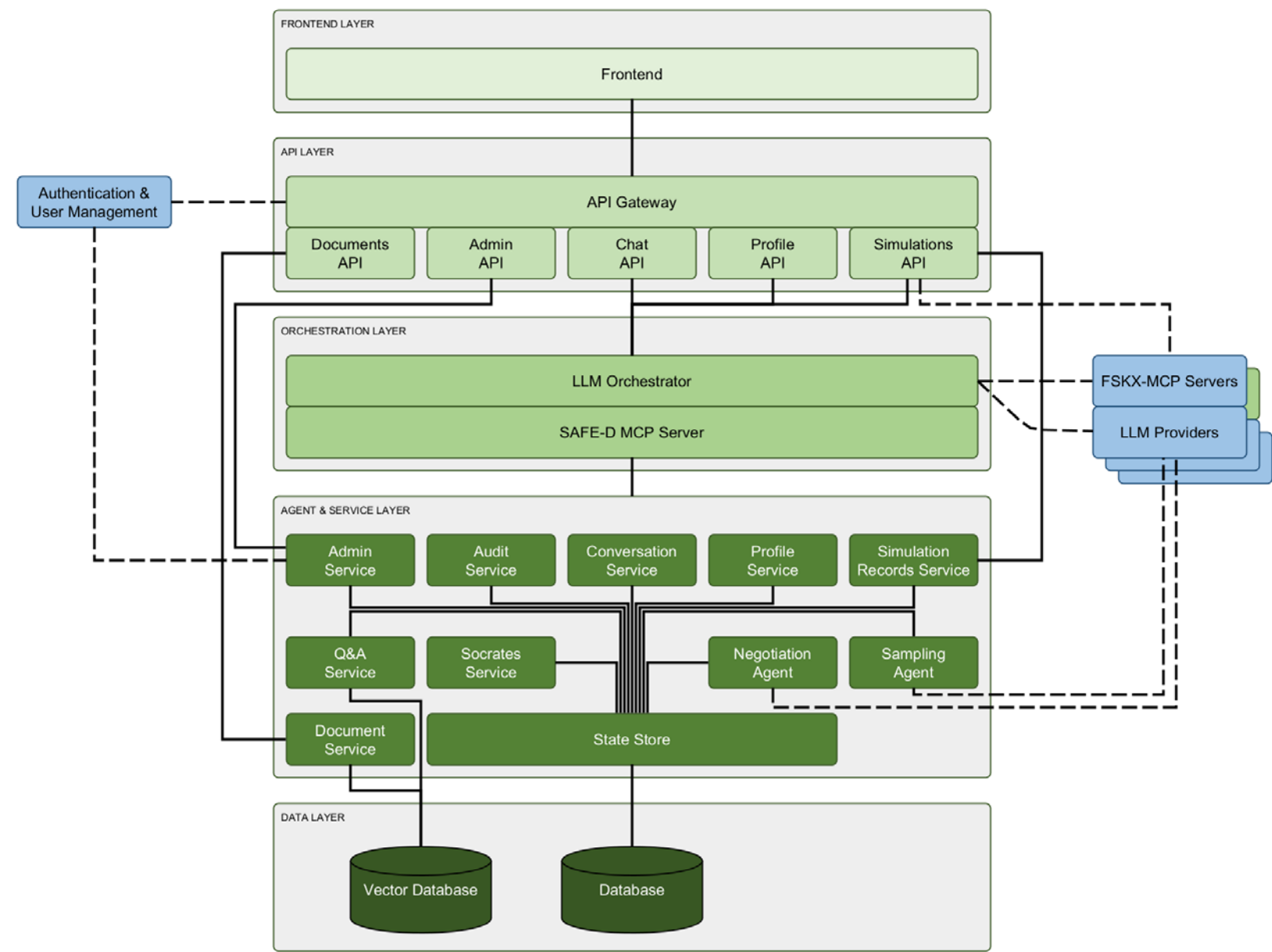
Confidence levels

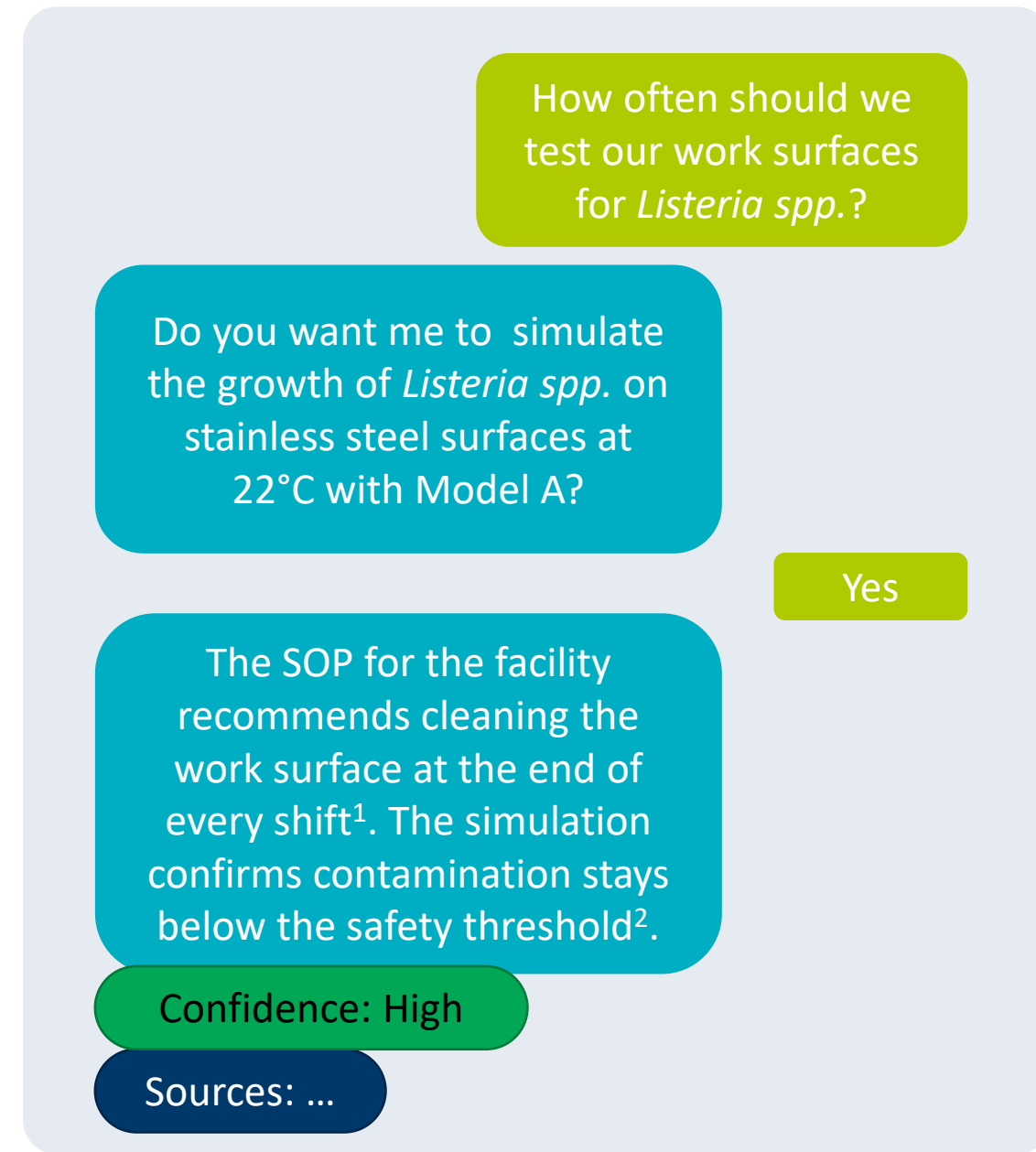
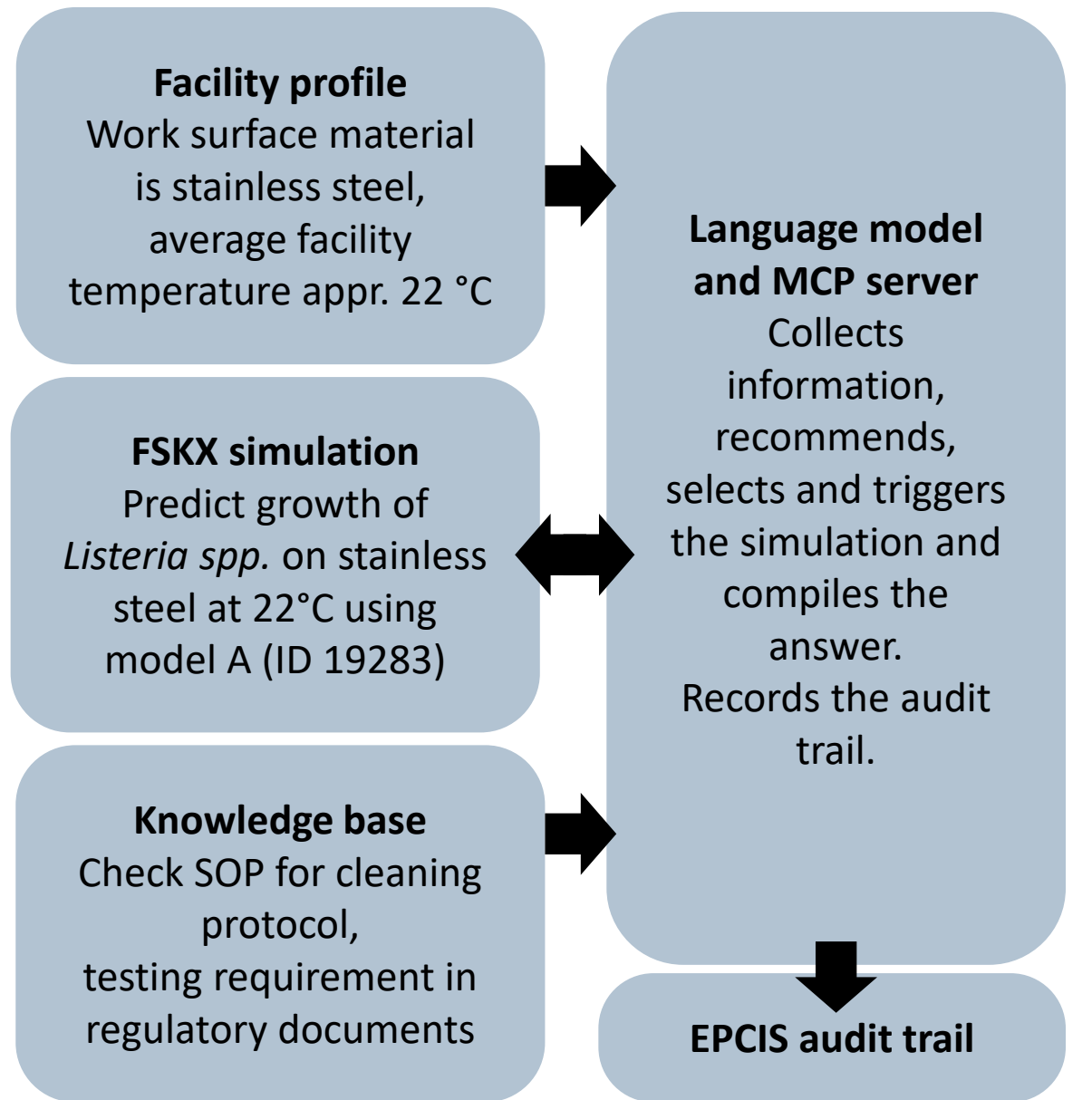
Explicit communication alongside every response indicates the need for additional information.

EPCIS audit trail

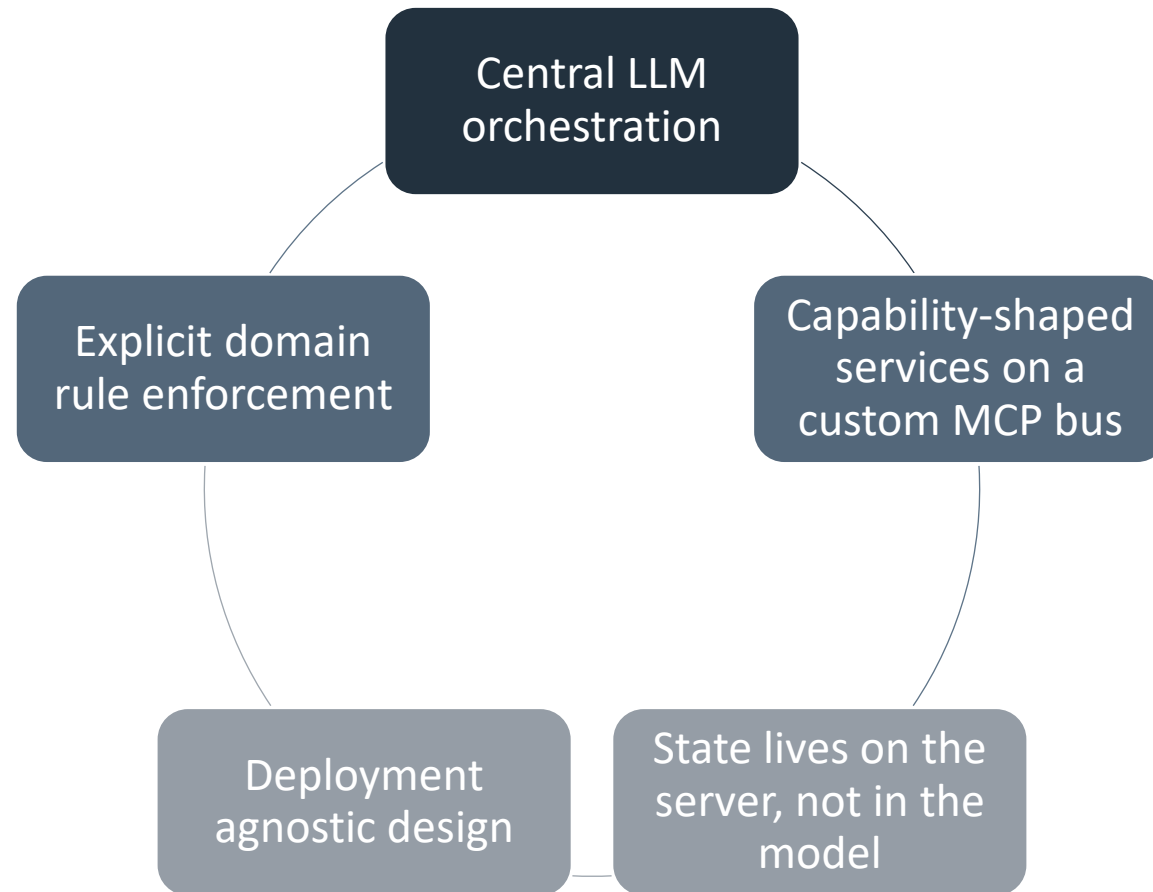
Each tool call is written to an immutable log to allow for decision reconstruction.

SAFE-D architecture





Design decisions



Known constraints and next steps

The LLM is load bearing

Which model can carry the weight is an empirical question, not a specification.

Token cost is elemental

Property of the design, not only tuning.
Prompt compression might be needed.

Towards real-world validation

Build on real-world requirements, the architecture needs validation as an implemented system.

RAG alone might not be enough

Exploring more elaborated solution like Knowledge Graphs or GraphRAG could elevate the architecture further.

Key takeaway

Building on modern standards like EPCIS, MCP and FSKX empowers a new generation of trustworthy AI beyond the limitations of LLMs.



Stefan Dehm

T +49 30 18412-51014

stefan.dehm@bfr.bund.de

German Federal Institute for Risk Assessment

bfr.bund.de/en



valid for texts produced by the BfR
images/photos/graphics are excluded unless otherwise indicated

BfR | Identifying Risks –
Protecting Health

Consumer health protection to go

BfR2GO – the BfR Science Magazine

bfr.bund.de/en/publications/bfr2go/

Follow us

 @bfrde | @bfren | @Bf3R_centre

 @bfrde

 youtube.com/@bfr_bund

 social.bund.de/@bfr

 linkedin.com/company/bundesinstitut-f-r-risikobewertung

 podcast.bfr.bund.de

 threads.net/@bfrde

 bsky.app/profile/bfrde.bsky.social