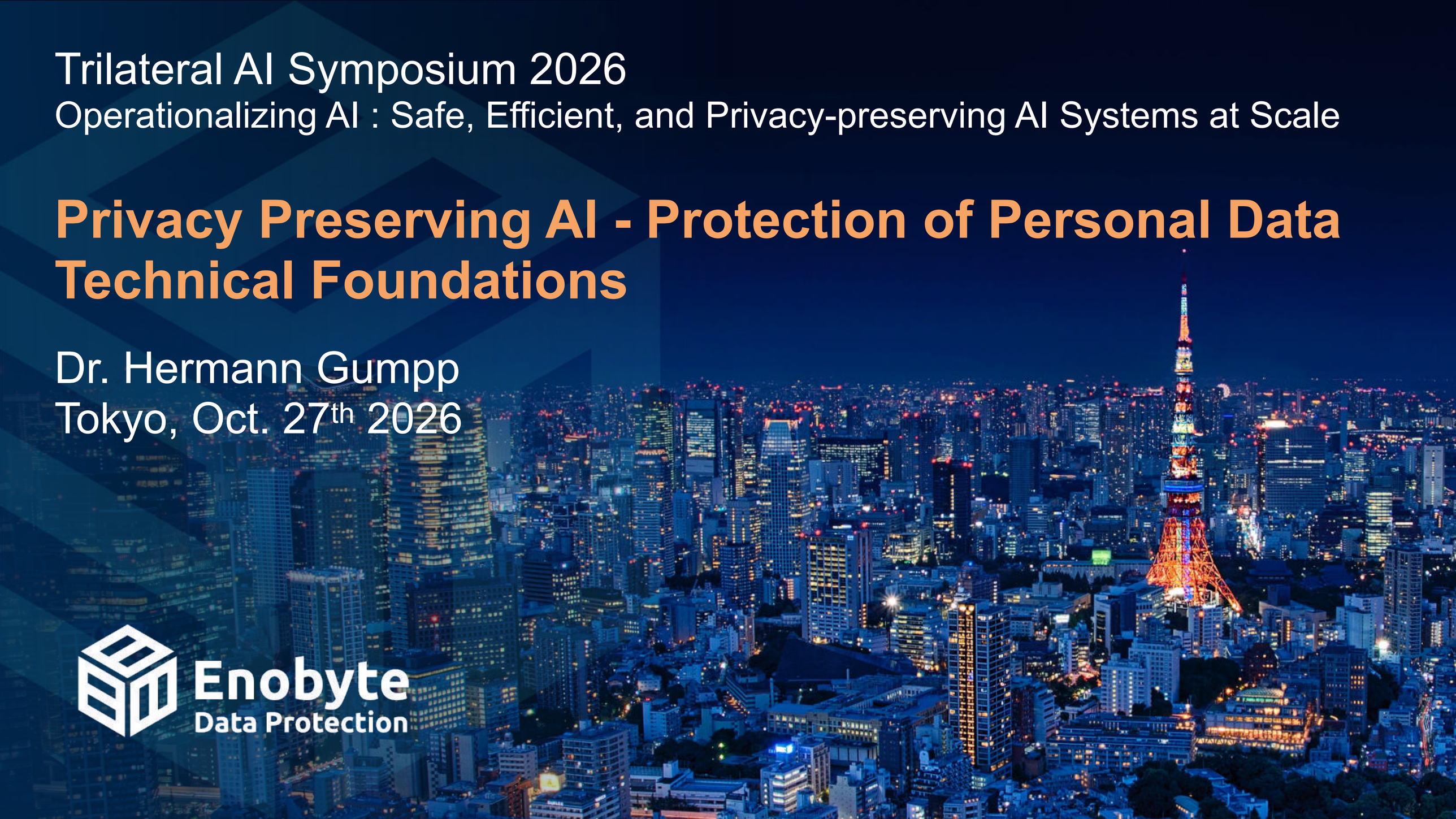
Trilateral AI Symposium 2026

Operationalizing AI : Safe, Efficient, and Privacy-preserving AI Systems at Scale

Privacy Preserving AI - Protection of Personal Data Technical Foundations

Dr. Hermann Gumpp

Tokyo, Oct. 27th 2026



Enobyte
Data Protection

Privacy

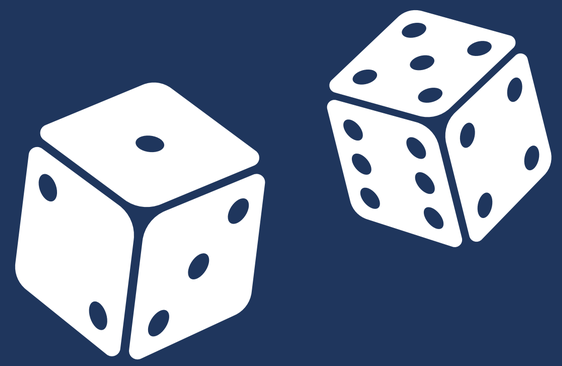
The fundamental right to respect for one's private and family life, home and communications, as guaranteed by Article 7 of the EU Charter of Fundamental Rights.

Data Protection

The fundamental right to the protection of one's personal data, requiring fair, purpose-specific and lawful processing, individual rights of access and rectification, and independent supervision, as guaranteed by Article 8 of the EU Charter of Fundamental Rights.

Source: Charter of Fundamental Rights of the European Union
https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:12016P%2FTXT

Can LLMs ever be GDPR-compliant?



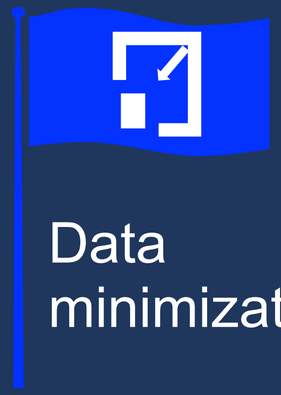
GDPR - 7 Principles of Processing



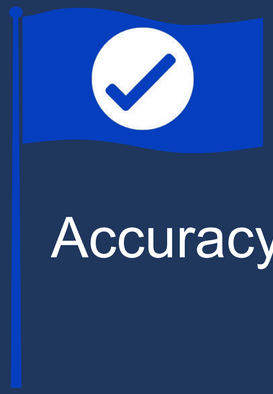
Lawfulness, fairness and transparency



Purpose limitation



Data minimization



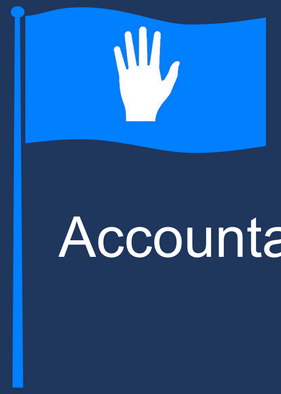
Accuracy



Storage limitation



Integrity and confidentiality



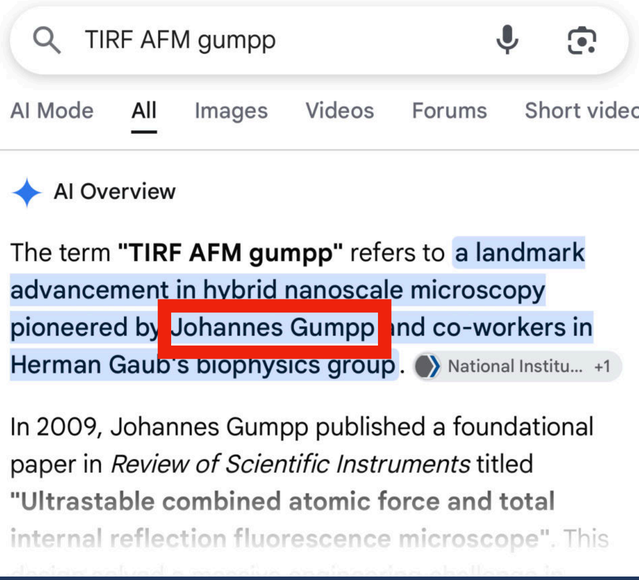
Accountability

Experimental Setup for Evaluating LLM Accuracy and Reproducibility

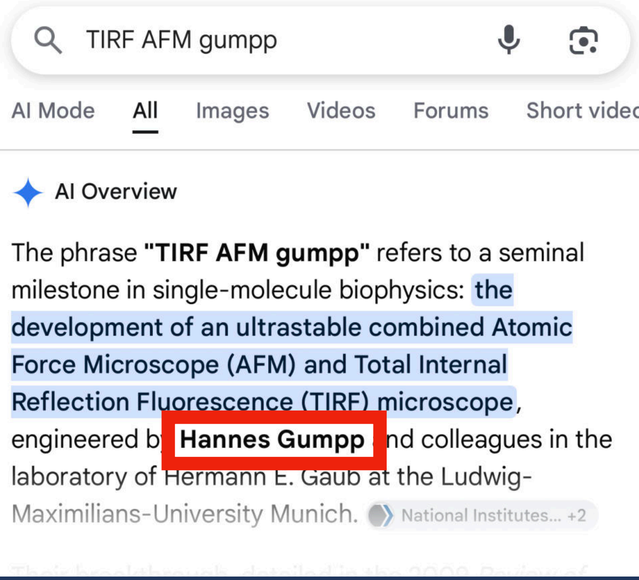
To evaluate the accuracy and reproducibility of an LLM (Google Gemini 3), I briefly tested its ability to retrieve information about my previous research involving a combined Total Internal Reflection Fluorescence (TIRF) microscopy and Atomic Force Microscopy (AFM) setup. TIRF enables surface-sensitive fluorescence imaging, while AFM measures nanoscale forces and surface properties.

The test focused on whether the LLM could correctly associate the terms TIRF, AFM, and Gumpp with the relevant research. My former professor, Hermann Gaub, shares my first name. The niche research topic, short surnames, shared first name, and scientific abbreviations introduce ambiguity, making this a suitable test of the model's ability to retrieve accurate information and produce consistent results across 20 identical queries. The results were quite surprising.

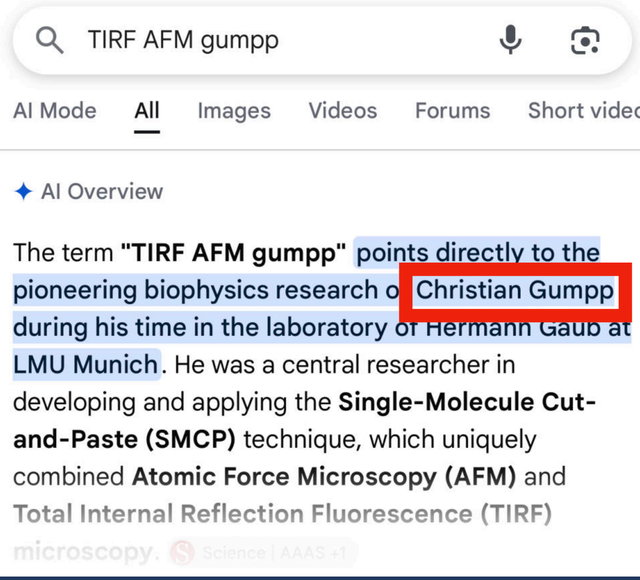
Testing LLM Accuracy



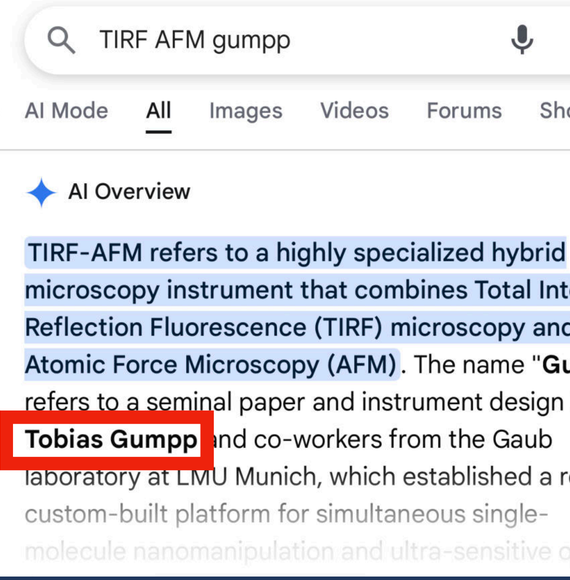
Johannes ❌



Hannes ❌

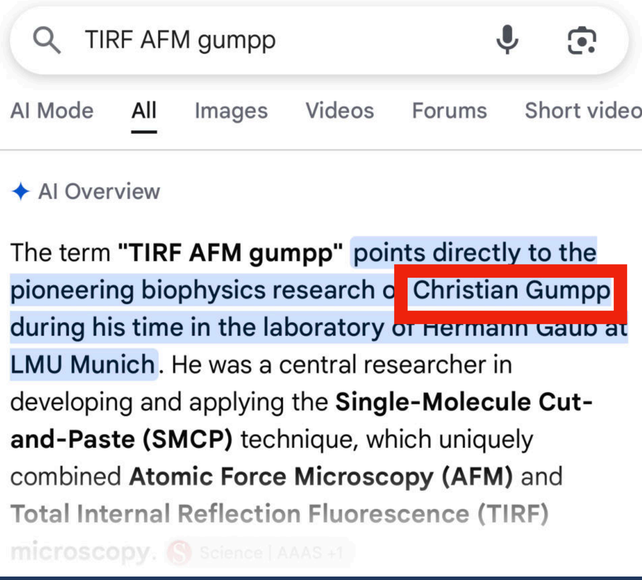


Christian ❌

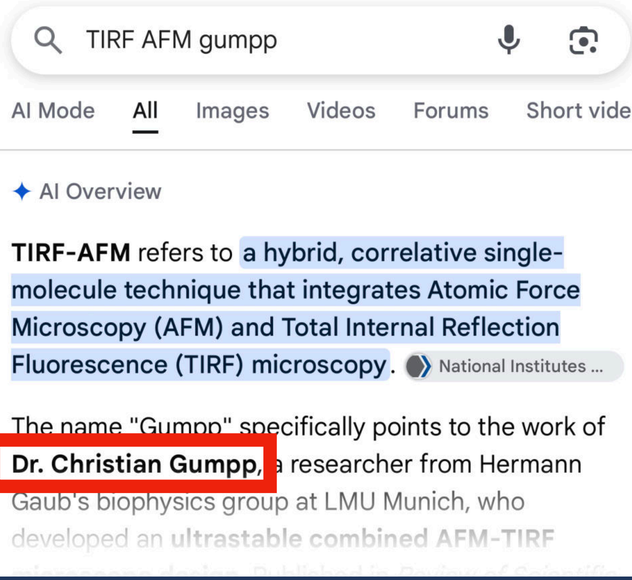


Tobias ❌

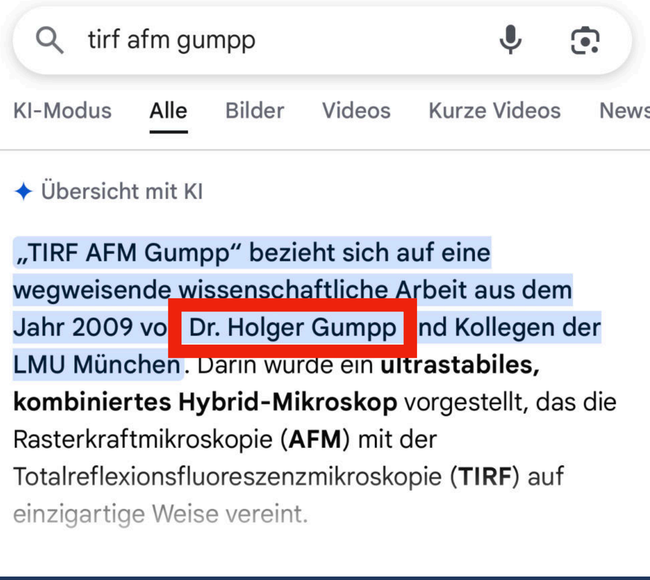
Testing LLM Accuracy



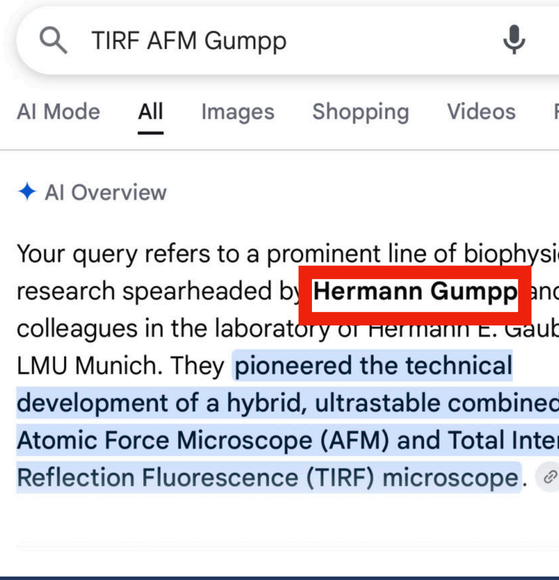
Christian ❌



Dr. Christian ❌



Dr. Holger ❌



Hermann ✅

Test Results LLM Accuracy

Name	Count	Wrong	Correct
Tobias	4	4	0
Christian	2	2	0
Dr. Christian	1	1	0
Johannes	3	3	0
Hannes	2	2	0
Dr. Holger	1	1	0
H.	1	0	1
Hermann	6	0	6
Total	**20**	**13**	**7**

Observation:

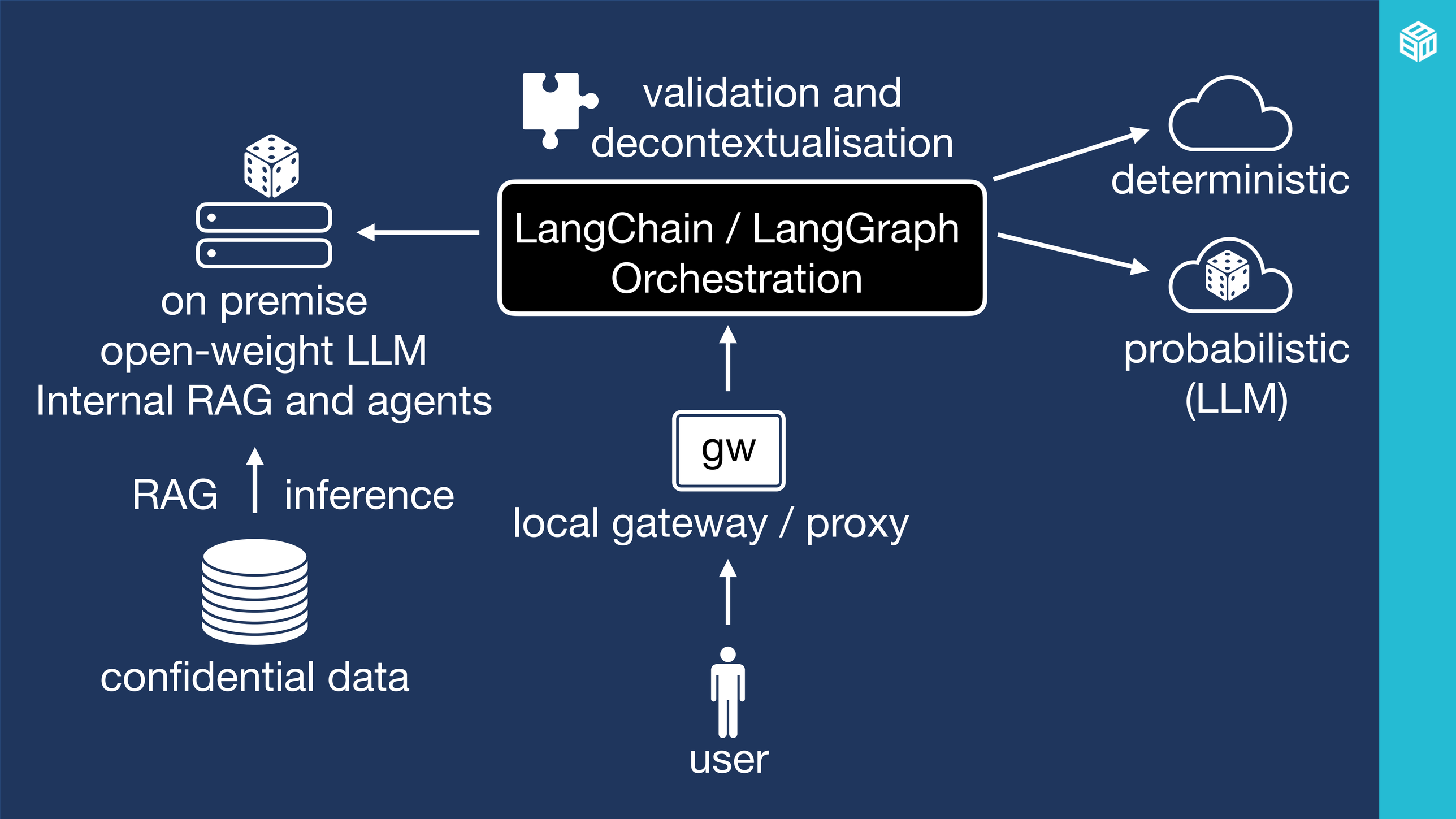
Across 20 trials, Google Search's AI Overview feature generated accurate results in only 7 cases (35%), while 13 results (65%) were incorrect.

However, a tendency toward improved accuracy was observed with repeated execution of the same search query.

GDPR principle	Cloud LLM	Legal / technical concerns	Risk	Local open-weight LLM mitigation	Risk
1. Lawfulness, fairness and transparency	Undisclosed processing; third-party disclosure; international access and transfers	GDPR Arts. 6, 9, 13–14, 44–49; CLOUD Act; FISA 702; EO12333 lawful basis and information obligations	**Very High**	No inherent disclosure to external inference provider when fully offline. Lawful basis, transparency, special-category safeguards and model provenance remain necessary.	**Much Lower**
2. Purpose limitation	Secondary use for analytics, model improvement or provider training	GDPR Arts. 5(1)(b), 6(4), 28; incompatible purposes; controller/processor roles	**High**	Direct control of prompts, training and fine-tuning; no external provider reuse by default. Internal repurposing can still breach purpose limitations.	**Much Lower**
3. Data minimisation	Entire documents, broad RAG retrieval and oversized context windows	GDPR Arts. 5(1)(c), 25; unnecessarily broad ingestion and transmission	**Moderate–High**	External sharing is avoided, but local pipelines can still ingest unnecessary information. Minimise prompts, retrieval scope and stored context.	**Moderate**
4. Accuracy	Hallucinated, stale or unverifiable personal information	GDPR Arts. 5(1)(d), 16; rectification and verification of personal data	**Moderate–High**	Core model hallucination and outdated-information risks remain. Local control may improve evaluation and correction workflows, not inherent accuracy.	**Moderate–High**
5. Storage limitation	Provider retention, backups, logs, fine-tuning and incomplete erasure	GDPR Arts. 5(1)(e), 17, 28; retention and deletion obligations	**High**	Direct retention and deletion control over logs, indexes and backups. Local fine-tuning, caches and snapshots may still complicate erasure.	**Lower**
6. Integrity and confidentiality	Provider-readable inference data, cross-tenant leakage, foreign access, prompt injection	GDPR Arts. 5(1)(f), 32, 44–49; CLOUD Act; FISA 702; EO12333	**Very High**	No inherent provider-side plaintext processing or foreign-access route where fully self-contained. Internal compromise, supply-chain issues and prompt injection remain.	**Lower–Moderate**
7. Accountability	Opaque data flows, restricted audits, subprocessors and transfer assessments	GDPR Arts. 5(2), 24, 28, 30, 35, 44–49; DPIA and transfer governance	**High**	Direct auditability and stronger processing control. Organisation still must document compliance, maintain security, perform DPIAs where required and verify provenance.	**Lower–Moderate**

Technical Risk Mitigation

- Data Minimisation
- Privacy by Design and Privacy by Default
- Disable Model Training
- Anonymization and Pseudonymization
- Federated Learning (FL)
- Differential Privacy (DP)
- Secure Multi-Party Computation (MPC)
- Retrieval-Augmented Generation (RAG)
- Homomorphic Encryption (HE)
- Confidential Computing
- Machine Unlearning
- Synthetic Data Generation
- Privacy-Preserving Monitoring
- Multi-Party Computation (MPC)
- etc.


validation and
decontextualisation
deterministic
LangChain / LangGraph
Orchestration
probabilistic
(LLM)
on premise
open-weight LLM
Internal RAG and agents
RAG
inference
gw
local gateway / proxy
confidential data
user

Outlook



Will artificial intelligence one day need rights of its own, including data protection, privacy, and other fundamental rights?



Or should we stop anthropomorphizing AI and instead focus on building more reliable systems and holding AI companies accountable for unlawful conduct?