



Connect

Jenseits des Hypes

Praktische SLM-Implementierungen für Enterprise GenAI

18. November 2025

Darmstadt





Stefan Bergstein
Senior Principal Chief Architect



Markus Gersdorf
Consultant



Steffen Röcker
Senior Account Solution Architect

Importance of small, specialized AI models

“The ability to have smaller, specialized models that you can train and run efficiently that’s I think where we will tap into a lot of value.”^[1]”

[1] [Red Hat CEO discusses importance of small, specialized AI models](#)



The Why

- Why small, specialized AI models?
- Use case across industries

The How

- NuExtract 1.5 tiny
- Event driven
- OpenShift AI

The Action

- Notebook
- Runtime
- Key takeaways





Weitere Anwendungsfälle

Healthcare

- **Medical Notes:** Doctors' notes often include diagnoses, treatments, and observations written in free text.
- **Radiology Reports:** Narrative descriptions of imaging results can be structured into diagnoses, measurements, and recommendations.
- **Patient Feedback:** Comments from patient satisfaction surveys can be analyzed for sentiment, common complaints, and areas for improvement.

Finance

- **Earnings Call Transcripts:** Statements and discussions during earnings calls can be converted into structured data for sentiment analysis and trend identification.
- **Financial News Articles:** Extracting stock movements, economic indicators, or corporate actions from free-text news articles.
- **Customer Queries:** Customer support chat logs or email inquiries to identify service issues or fraud patterns.

Retail

- **Customer Reviews:** Free-text reviews can be analyzed to extract sentiment, key features, or areas of concern.
- **Social Media Mentions:** Tweets, posts, and comments about brands or products can be analyzed for brand reputation and marketing insights.
- **Product Descriptions:** Unstructured product specifications can be converted into structured formats for better catalog management.

Education

- **Student Feedback:** Free-text feedback from course evaluations can help identify areas for curriculum improvement.
- **Research Papers:** Summarizing key findings, extracting references, or categorizing papers by topic.
- **Online Forum Discussions:** Identifying common questions or challenges faced by students.

Legal

- **Case Law and Judgments:** Court opinions and rulings can be structured into precedents, legal principles, and key entities.
- **Contracts:** Extracting clauses, obligations, deadlines, and parties from lengthy legal documents.
- **Regulatory Texts:** Structuring compliance requirements from regulations and policies.

Technology

- **Bug Reports:** User-submitted bug descriptions can be structured to identify patterns and prioritize fixes.
- **Forum Discussions:** Conversations on technical forums can reveal trends in customer issues or feature requests.
- **Logs:** System logs, which often have a mix of structured and unstructured data, can be mined for errors, patterns, or anomalies.

Media and Entertainment

- **Movie or TV Show Reviews:** Audience reviews can be analyzed for common themes, preferences, and critiques.
- **Script Analysis:** Extracting character names, dialogues, or scene descriptions from screenplays.
- **Social Media Trends:** Identifying trending topics or user sentiment from platforms like Twitter or Instagram.

Transportation and Logistics

- **Driver Logs:** Unstructured logs describing routes, stops, and delays can be structured for performance analytics.
- **Customer Complaints:** Emails or feedback forms with details about delays, damages, or service quality.
- **Weather Reports:** Structuring data from narrative forecasts for route optimization.

Energy and Utilities

- **Inspection Reports:** Field reports written by technicians about equipment performance or repairs.
- **Customer Inquiries:** Emails or chat logs regarding billing, outages, or service issues.
- **Incident Reports:** Unstructured descriptions of accidents or system failures.

Government and Public Sector

- **Citizen Feedback:** Free-text submissions to government platforms for issues like potholes, public safety, or municipal services.
- **Legislation and Policy Documents:** Structuring laws into clauses, responsible agencies, and timelines.
- **Incident Reports:** Narrative descriptions of events like crimes or natural disasters for emergency response planning.

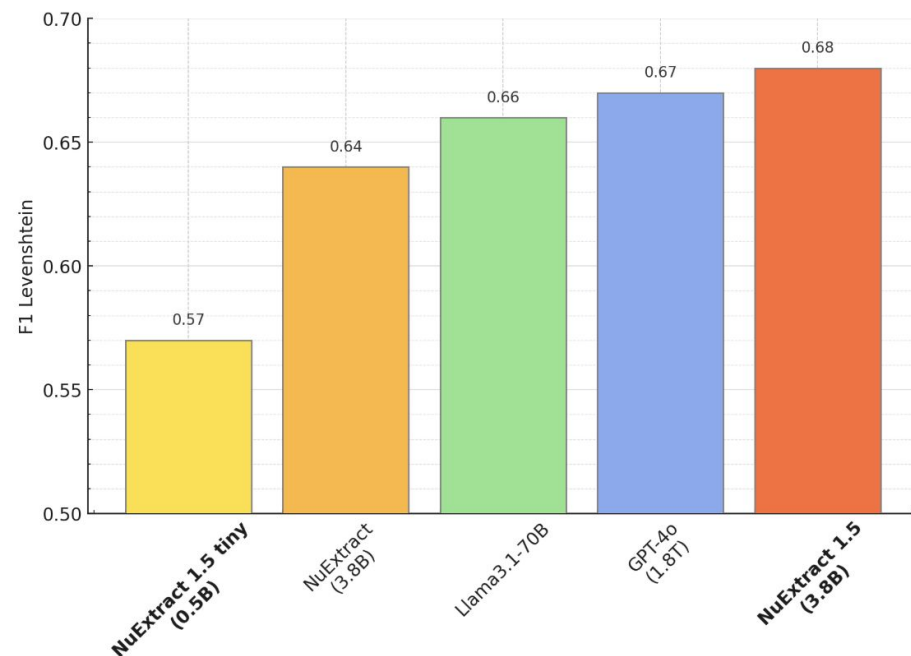
Insurance

- **Claims Descriptions:** Narratives provided during claim filings can be structured into cause, extent of damage, and claim amount.
- **Adjuster Notes:** Field adjusters' notes on inspections can be structured for claim processing.
- **Policy Documents:** Extracting terms, coverage details, and exclusions from policies.



Structured Extraction with NuExtract 1.5 tiny

NuExtract-tiny-v1.5 is a fine-tuning of Qwen/Qwen2.5-0.5B, trained on a private high-quality dataset for structured information extraction.



Zero-shot results on the structured extraction benchmark. Average F1-score over extraction problems. NuExtract 1.5 is better than NuExtract and slightly better than GPT-4o.

Lot de 10 assiettes Happy - fuchsia, en carton, mesurant 22,5 cm de diamètre.

Assiettes anniversaire - vert anis

Pour l'occasion, nous vous proposons, ici, un lot de 10 assiettes Happy, couleur fuchsia, en carton. Ces assiettes illustrées du mot Happy en doré, mesurent 22,5 cm de diamètre

Input text

```
{
  Product Information: {
    Product Name: ,
    Color: ,
    Material: ,
    Size: ,
    Quantity:
  },
  Description: ,
  Related Products: [
    {
      Product Name: ,
      Quantity:
    }
  ]
}
```

template

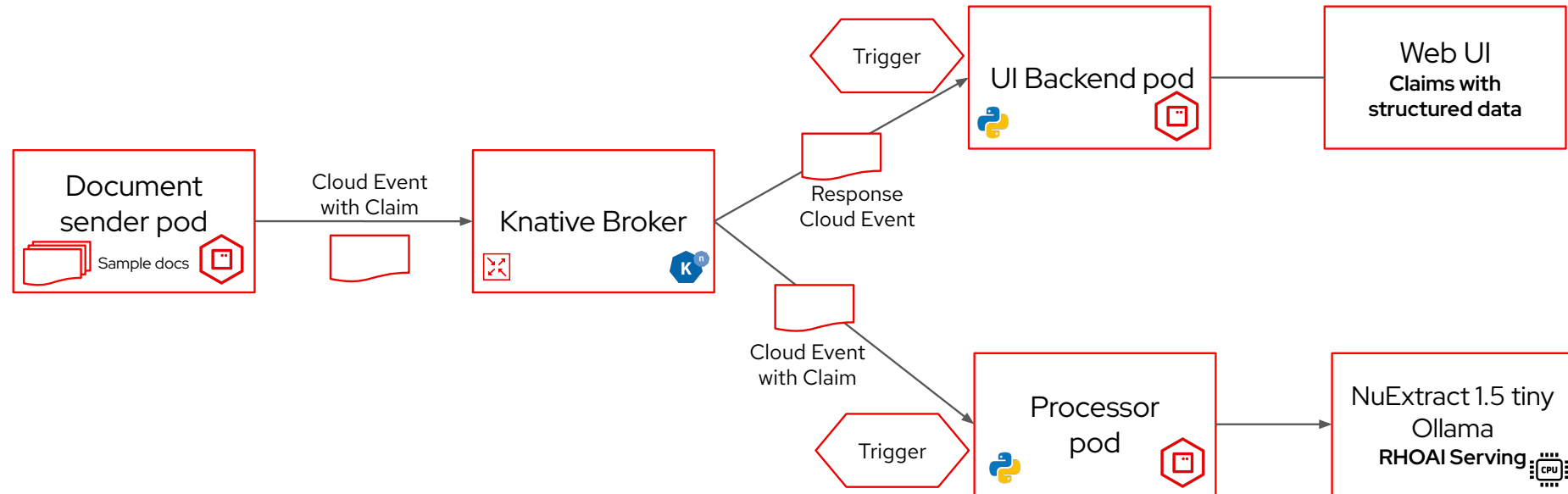
```
{
  Product Information: {
    Product Name: assiettes Happy ,
    Color: fuchsia,
    Material: carton,
    Size: 22,5 cm,
    Quantity: 10
  },
  Related Products: [
    {
      Product Name: assiettes anniversaire - vert ani,
      Quantity:
    }
  ]
}
```

output

Training example with a French document and an English template.



Demo Architektur



Hands-on Lab:

<https://github.com/happy-octo/jenseits-des-hypes>



OpenShift Login:

Username: user2 - user48

Password: hello

MinIO:

Username: minio

Password: minio123



Key Takeaways

1. Unstructured text streams are common across various industries and often contain valuable information that can be leveraged for decision-making, analytics, or operational efficiency.
2. Extracting structured information from these streams can be highly beneficial.
3. The demo showcases structured information extraction with small language models (SLM).
4. SLM use less resources, can be deployed in edge environments, use less energy, and do not required powerful GPUs.

Attention CPU is all you need





Jetzt Session bewerten!

Einfach QR-Code scannen,
Session aus der Liste wählen
und bewerten. **Vielen Dank!**

red.ht/rhsc-darmstadt-feedback

Red Hat
Summit

Connect

Thank you



linkedin.com/company/red-hat



facebook.com/redhatinc



youtube.com/user/RedHatVideos



twitter.com/RedHat



Appendix



Semantic Sonnenschirm

TIME	NAME	ADDRESS	POLICY	TELEPHONE	ACCIDENT LOCATION	DATE	TIME	CLAIM	CLAIM DETAILS
2024-12-06 13:52:58	Emily Crawford	5173 Jones Manor Apt. 693, Paulville, HI 14002	YZ-941964228	200.204.6602x285	intersection of Maple Rd and Elm Ln	2024-01-05	approximately 4:00 PM	Dear Parasol Insurance, My name is Emily Crawford, and I am writing to file a c...	View Details
2024-12-06 13:52:23	Kimberly Hayden	Nancyborough, NJ 83953	BC-607767656	441.285.9885x9255	intersection of Birch St and Willow Ave.	2024-01-03	1:45 PM	Dear Parasol Insurance, My name is Kimberly Hayden, and I am writing to file a ...	View Details
2024-12-06 13:51:49	Tina Serrano	Youngport, IA 45769	XY-789752615	422-346-9787	Willow Ave and Oak St.	2024-01-05	3:30 PM	Dear Parasol Insurance, My name is Tina Serrano, and I am writing to file a cla...	View Details
2024-12-06 13:51:14	Ashley Williams	5315 Jason Course, North Carlachester, NH 80896	XY-266394153	(284)210-3760	Cedar Dr and Oak St.	2024-01-04	5:15 PM	Dear Parasol Insurance, My name is Ashley Williams, and I am writing to file a ...	View Details
2024-12-06 13:50:40	Jill Rodriguez	1557 Miranda Parkway Suite 056, New Calvinhaven, IN 19547	BC-590714306	888-017-5120x4262	intersection of Birch St and Main St.	2024-01-02	11:45 AM	Dear Parasol Insurance, My name is Jill Rodriguez, and I am writing to file a c...	View Details
2024-12-06 13:50:03	Anna Miller	4609 Williams Alley Apt. 216, South Kellymouth, AL 03858	AC-357458259		Spruce Ct and Main St.	2024-01-05	11:45 AM	Dear Parasol Insurance, My name is Anna Miller, and I am writing to file a clai...	View Details

