

# Measuring the Quality of GenAI Systems

Liang Yu



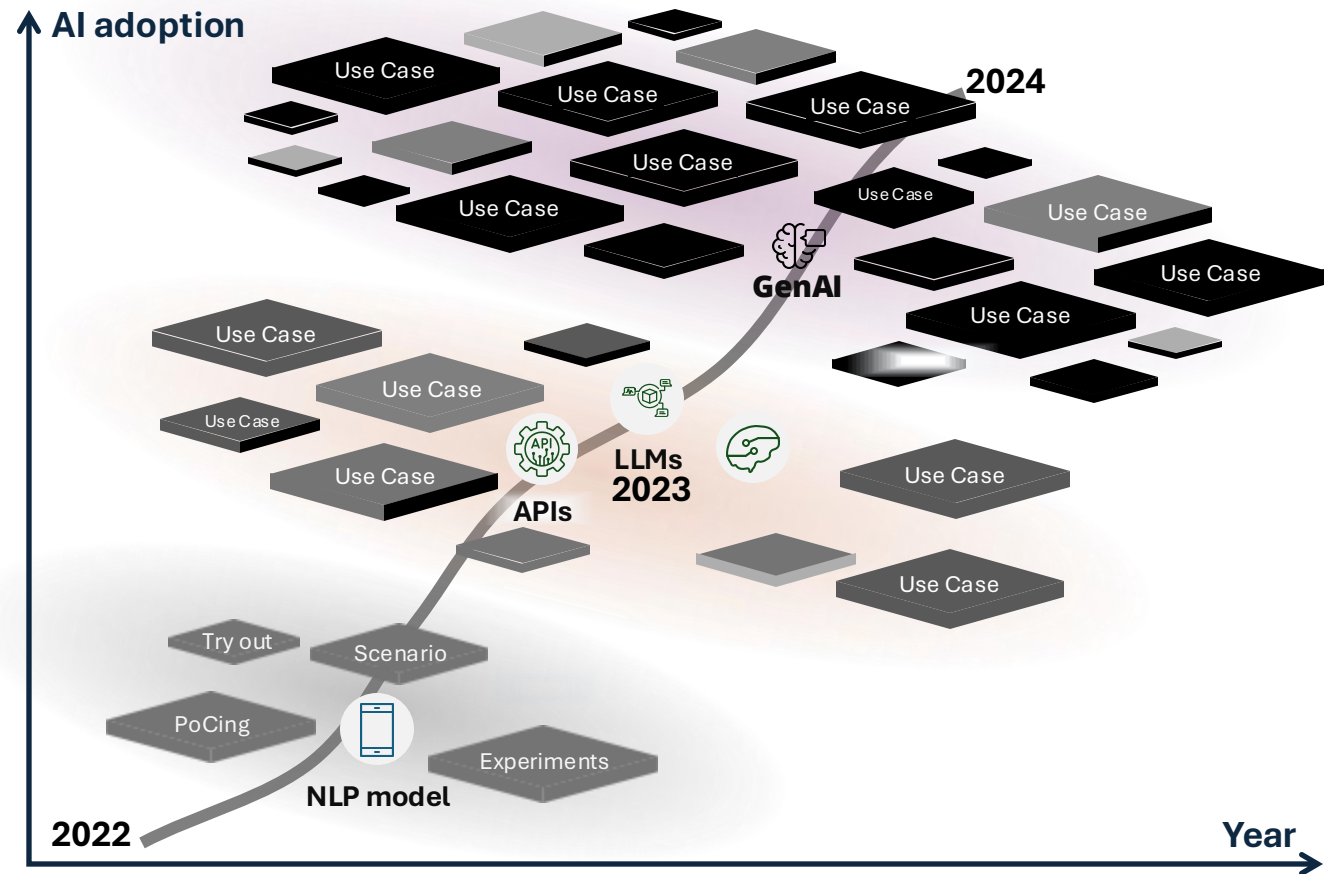
A developer works at Ericsson Mobile Financial Services (EMFS)

A researcher at BTH (50%)

GenAI: Generative Artificial Intelligence

# Our Journey

- Dec 2022 to present
- New use cases are emerging.
- How to measure those emerging use cases?

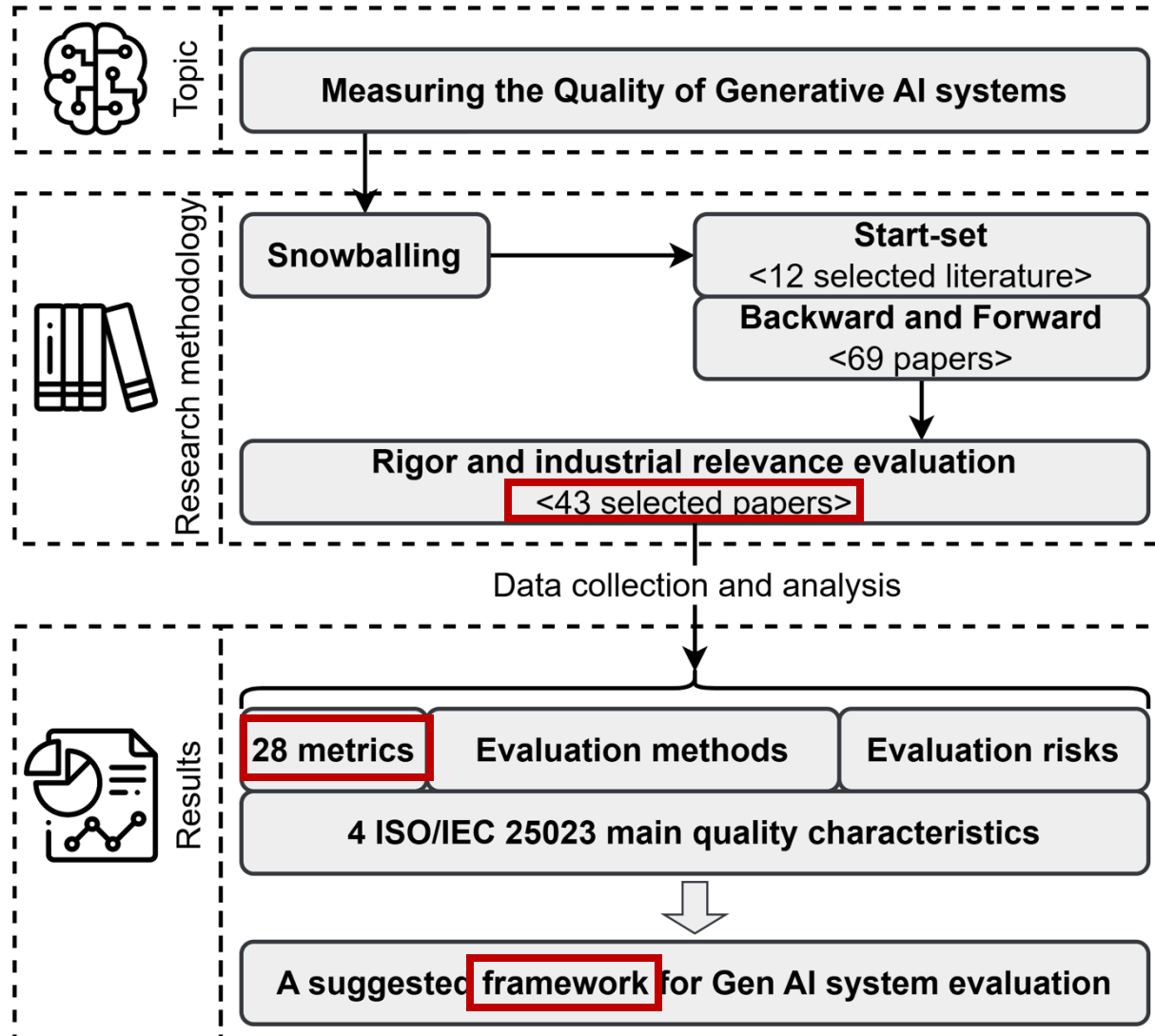


NLP: Natural Language Processing; LLM: Large Language Model

You can't manage what you can't  
measure.

– Peter Drucker

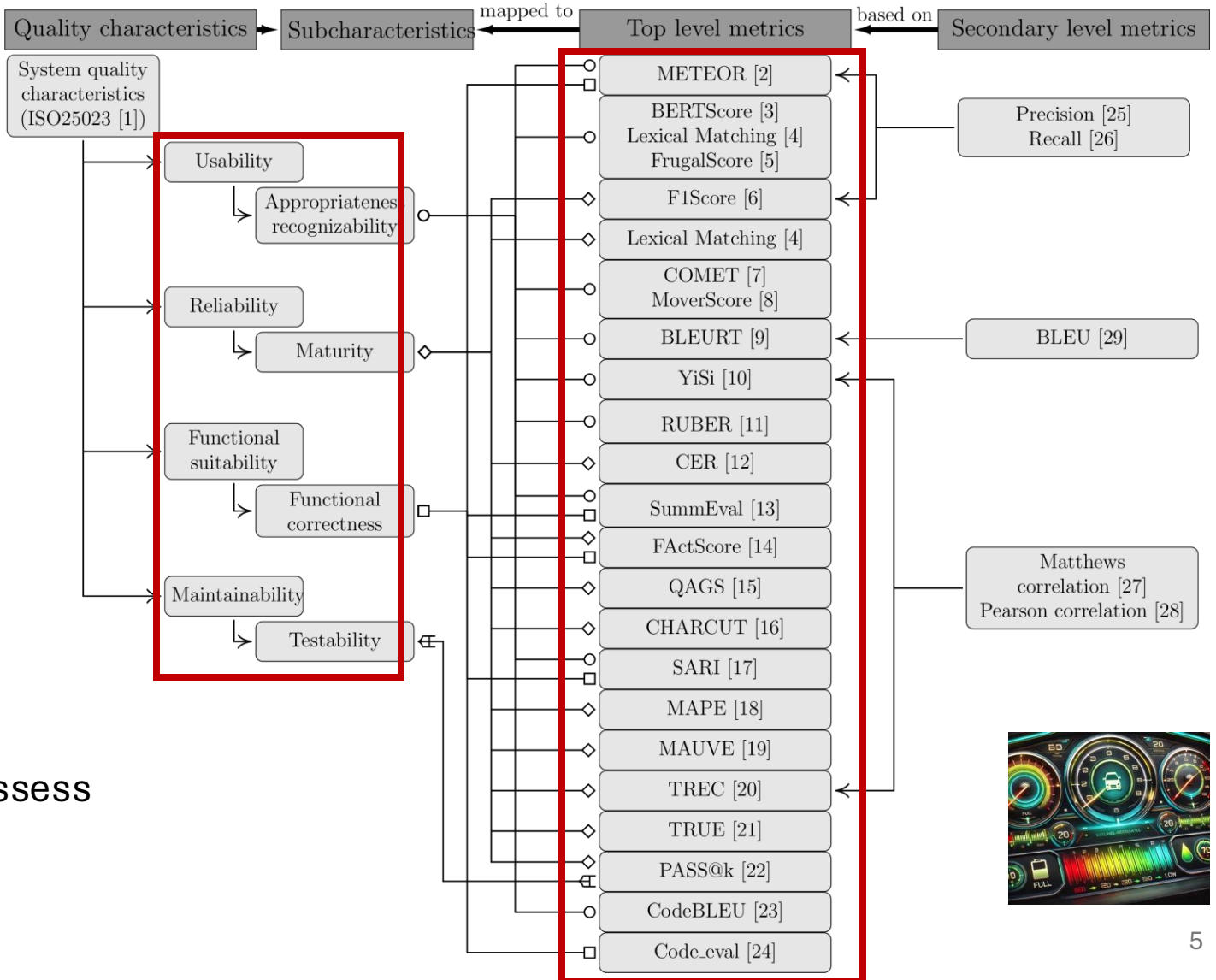
# Our research



Mapping metrics to ISO quality characteristics.

# Research finding (1/2)

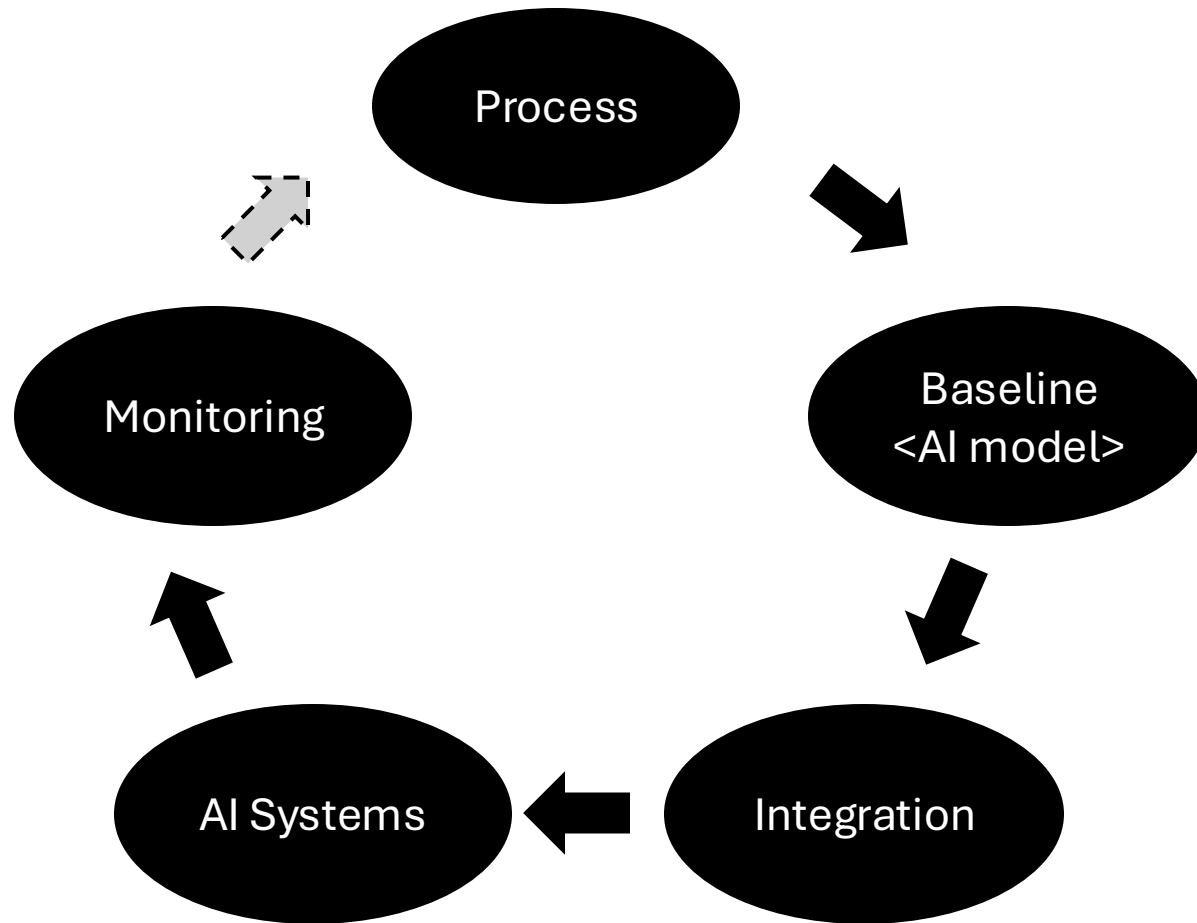
Metrics can be used to measure ISO quality characteristics.



Metric: a quantifiable measure to assess how well a system performs.



# Research finding (2/2)



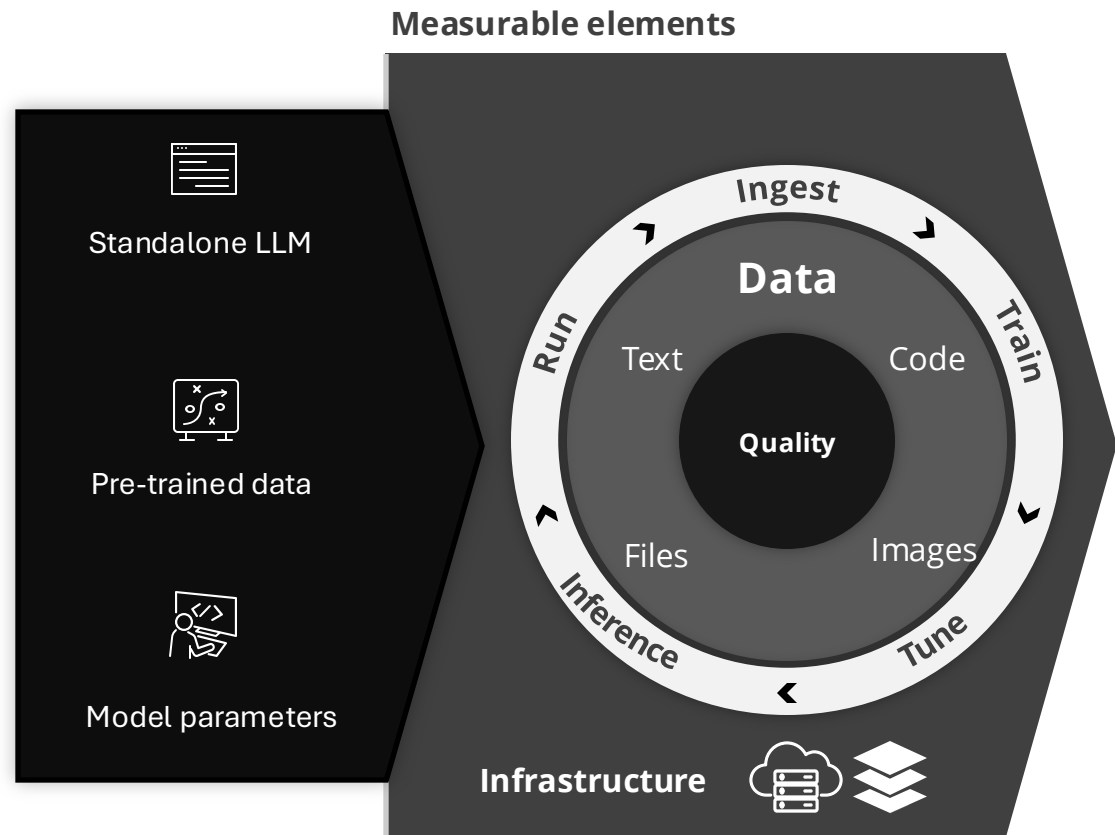
**Process:** Identify metrics and metric data.

**Baseline:** the foundation for evaluating LLM's outputs before implementing any adaptation or enhancements.

**AI System Integration:** Extend the baseline to enterprise datasets with industrial context.

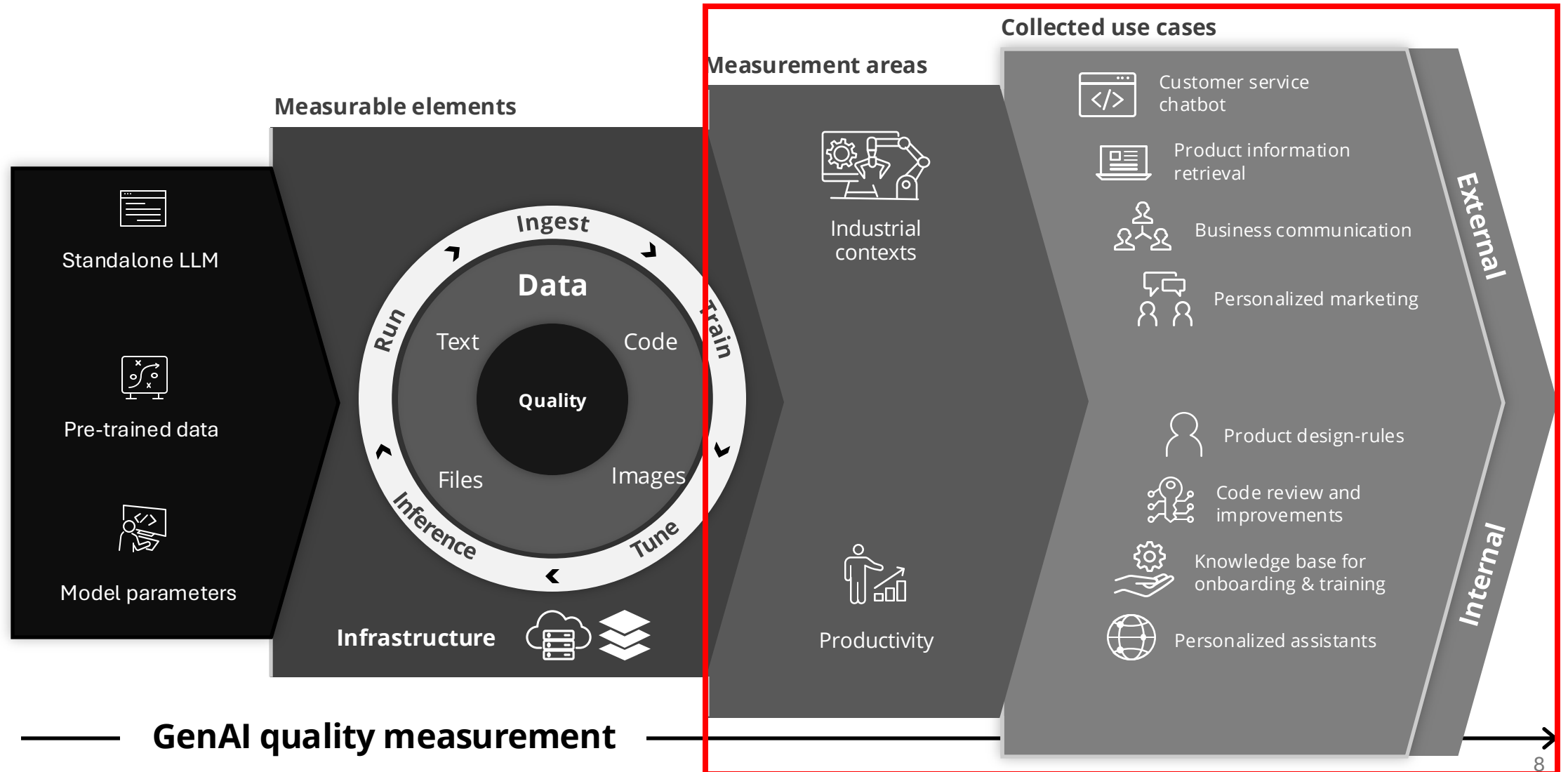
**Monitoring:** Dashboard with quantifiable metric scores to ensure the AI systems perform well.

# Industry - Data



GenAI quality measurement →

# Industry - use cases





# Industry - use case example

Description	Data source	Business impact
<p>Given the need of onboarding new team members, product information sharing becomes useful.</p>	<p>Texts:</p> <ul style="list-style-type: none"> <li>• Knowledge or information stored in folders, files, and scripts.</li> </ul>	<ul style="list-style-type: none"> <li>• Enable fast knowledge sharing.</li> <li>• Decrease the time required for information retrieval.</li> </ul>
Example	Metrics	
<p>Questions/answers chatbot for software product architecture, design rules, and implementation.</p>	<p>Response time</p> <p>BERTScore – Semantic similarity</p>	



Information Retrieval Bot

 Enable SocketsModel to use  Disable History

Send &gt;



# Takeaway



## Identify use cases

System qualities



## Start with small

Measurable use cases



## Check your data

Available enterprise data



## Build your AI talent

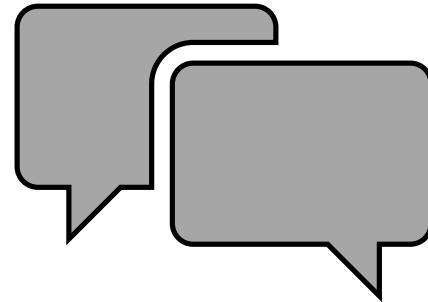
Skills your employees



## Continuous quality tracking

Metrics

# Let us discuss ...



Contact: Liang Yu  
<liang.yu@bth.se>