



SES108

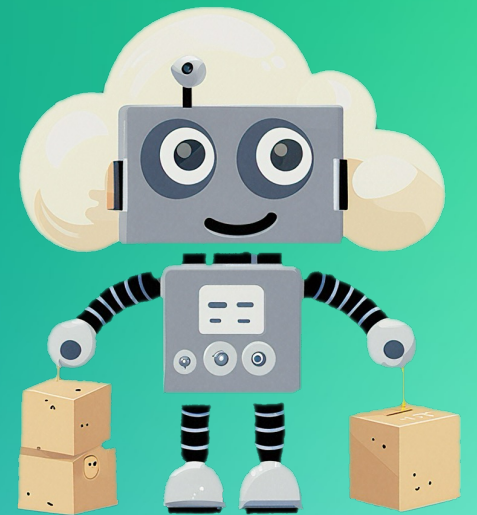
How Lounge by Zalando Innovates Content Localization with Generative AI

Gaiar Baimuratov

Head of Product and Tech
Lounge by Zalando

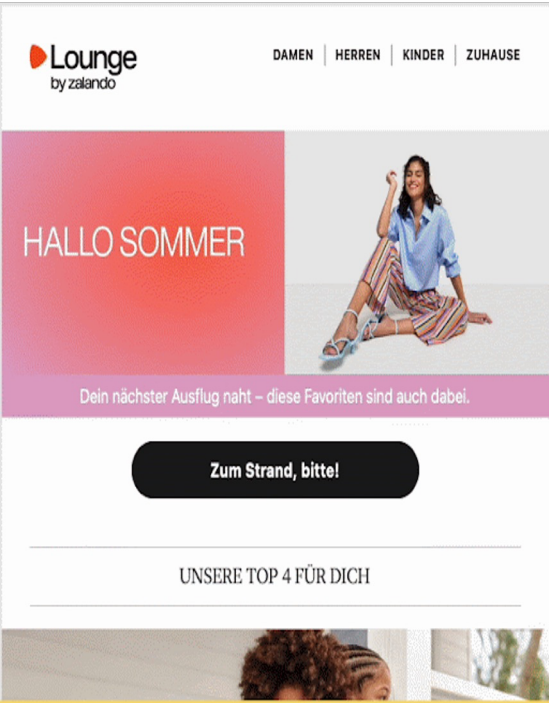
Nikita Kozodoi, PhD

Senior Applied Scientist
Amazon Web Services

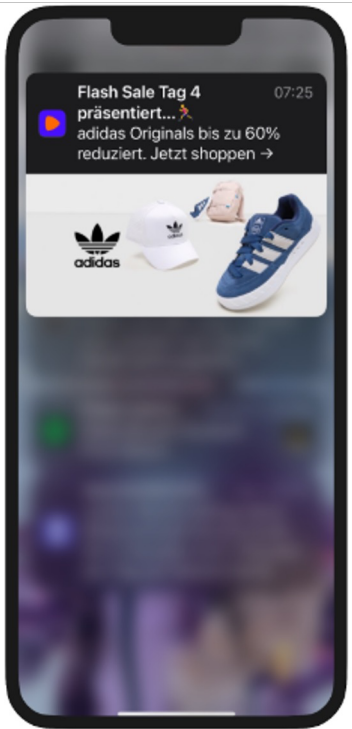


Connecting Through Content

Newsletter

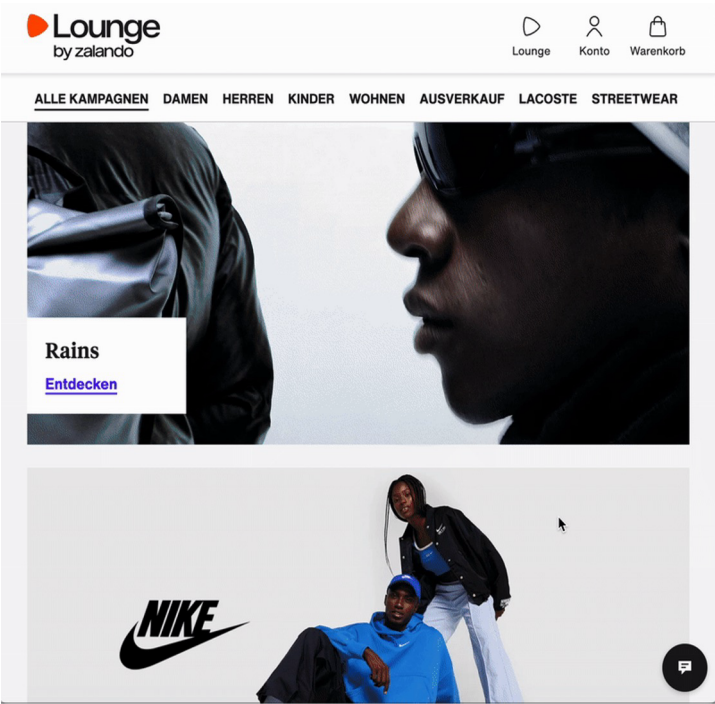


Push Notification



30% to 90% efficiency increase & cost savings

Product Detail Page



We are the leading
multi-brand fashion
and lifestyle destination
in Europe

About

15.3bn

Zalando GMV

>52m

Active Customers

19

Lounge Markets

Zalando

New Lounge Brand Positioning

The go-to-destination to **DISCOVER** curated deals in
QUALITY fashion and **LIFESTYLE**, every day.

A curated offering with
personalised experiences

An opportunity
for daily savings

New offers every single
day of the week

The Localization Journey: Bringing Campaigns to Life

Tempo di regali



Darčeková sezóna



Šventinis sezonas



Zeit für Geschenke



Cadeautjestijd



Localization effort for a tailored and meaningful customer experience

Concept

Copy

Localization x 20

Content Evaluation

Implementation
x 23 markets

Repeat

Customers Expect More



Relationship Building

- **Customers want to be seen**
- Strengthen brand perception
- Increase retention
- NPS & revenue uplift



Enhances Trust

- **80% of customers are more likely to purchase from brands offering personalized experience**
- Accurate, detailed & informative product descriptions
- Consistent messaging strengthens brand credibility



Adapts to local Markets & trends

- **71% of consumers expect tailored interactions**
- React swiftly to market changes
- Customers demand value-based and context-aware content

Challenges in Localized Content Creation

High volumes

+10k

Product descriptions
localized weekly

Resource intensive

+250h

Weekly hours spent on
content localization

Content evaluation

~15k€

To ensure SEO text
was accurate

Why we chose AWS

 Needs

- Localization at scale
- Speed to Market
- Operational efficiency
- Automated quality control
- Platform integration

**GenAI Innovation Center**

- Generative AI Innovation Center
- Machine Learning expertise
- Evaluation & revision expertise
- External Technical support
- Competitive pricing

Automated Localization with GenAI on 

1. Content Localization

2. Tonality Processing

3. New Content Ideas

Settings

Language model

How to use the app?

Tonality Guide

Select guides

Tonality guides:

AWS_CRM TONA... x

Select all tonality guides

Add custom instructions

Custom instructions:

Enter custom instructions

Examples, tonality tips, etc

Using 1 tonality guide(s)

Inputs



Generate Content

Revise Content

Evaluate Content

Parameters

Content type:

Leave empty to determine automatically

Number of self-revision rounds:

0 5

Output languages:

EN x DE x ES x FR x x

Select all languages

Translate one language at a time

Input file

Upload your file or provide field values in English:

Upload Excel file

Enter fields manually

Download example Excel format:

Excel template

! Excel file may have multiple sheets and tables. Each sheet must have 3 columns: **module**, **character limit**, **base copy**. See template above.

Upload your Excel file:



Drag and drop file here

Limit 200MB per file • XLSX

Browse files



Clear results



Create localizations

Settings

🧠 Language model ▾

❓ How to use the app? ▾

Tonality Guide

📁 Upload tonality guides ▾

📁 Select tonality guides ▾

Input File

Select sheet:

Daily Newsletter ▾



Generate Content

Revise Content

Evaluate Content

⚙️ Parameters

Content type: ?
Daily Newsletter ▾

Number of self-revisions: ?

0

5

Output languages: ?

DE × ES × FR × × ▾

☐ Select all languages

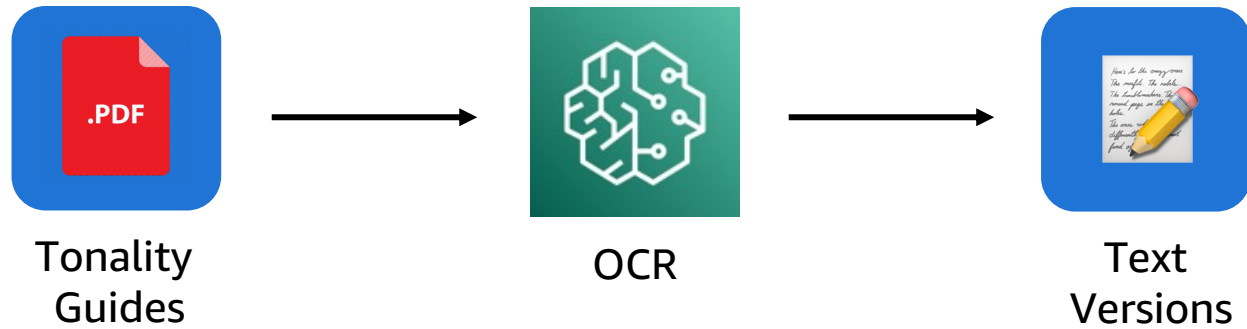
Inputs ▾

Clear results

Create localizations

Sign out

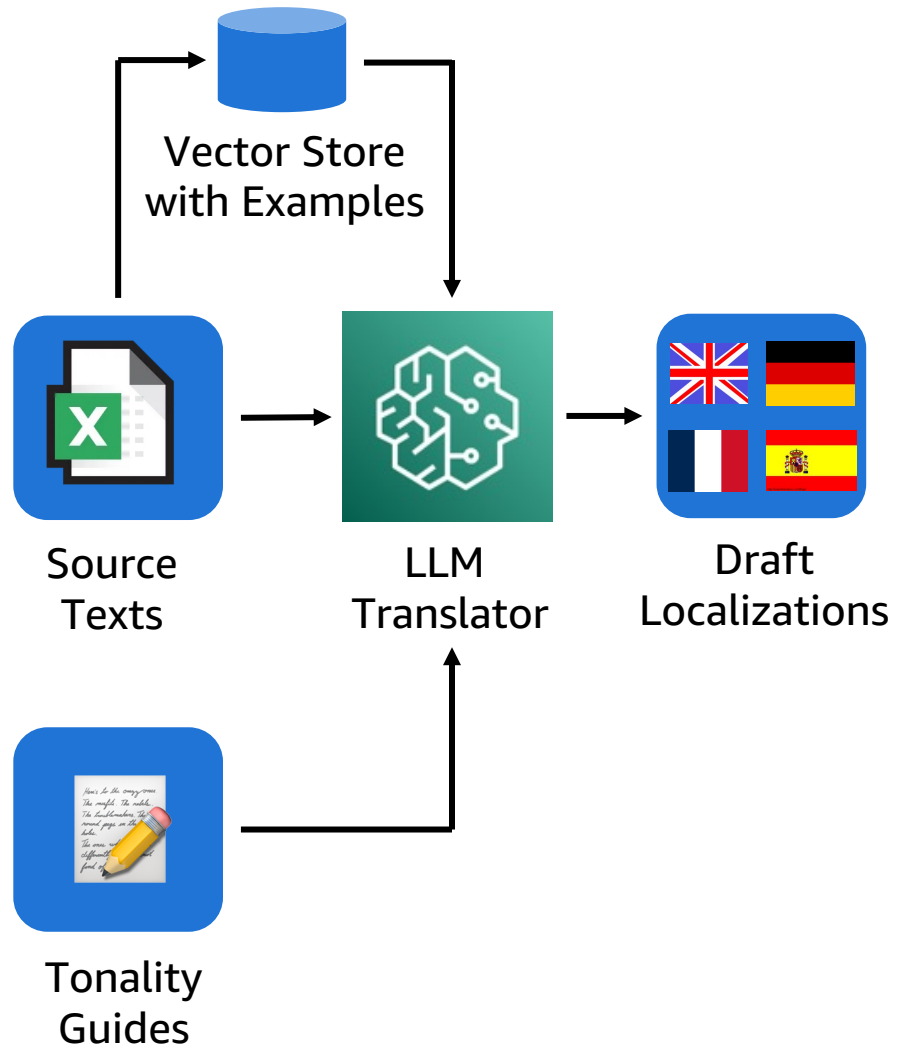
Generative AI Solution: Step 1



Parsing tonality guides

Documentation containing the style guide for different locations, branding guide, naming conventions, and other important information.

Generative AI Solution: Step 2

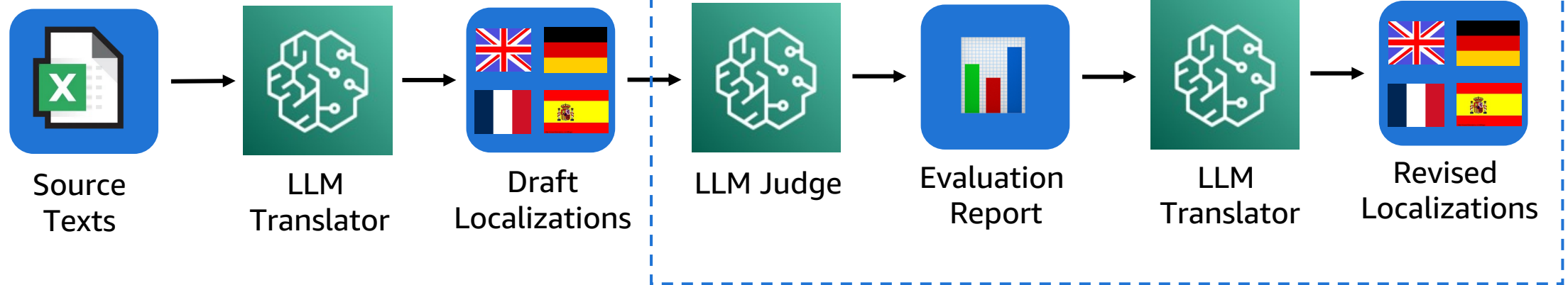


Generating localizations

Creating initial versions of the localized texts using the tonality guides and most relevant few-shot example translations as part of the LLM prompt.

Generative AI Solution: Step 3

Self-Reflection Module

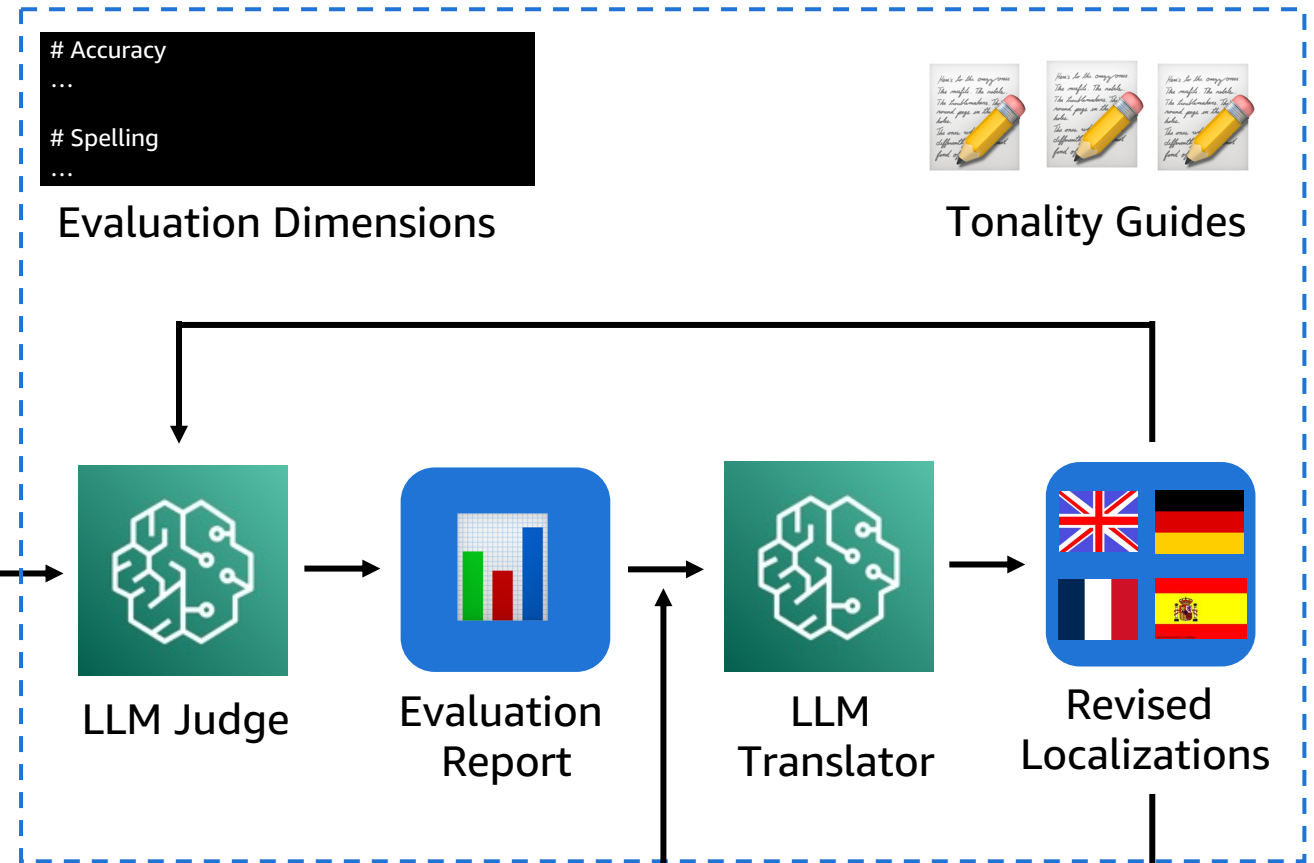


Iterative self-reflection

Multiple improvement iterations with evaluation and revision based on pre-defined evaluation criteria.

Generative AI Solution: Step 4

Self-Reflection Module



Human in the loop

Expert evaluation and refinement before finalizing.



Ingredient 1: Anatomy of a High-Quality Prompt

Persona and Context

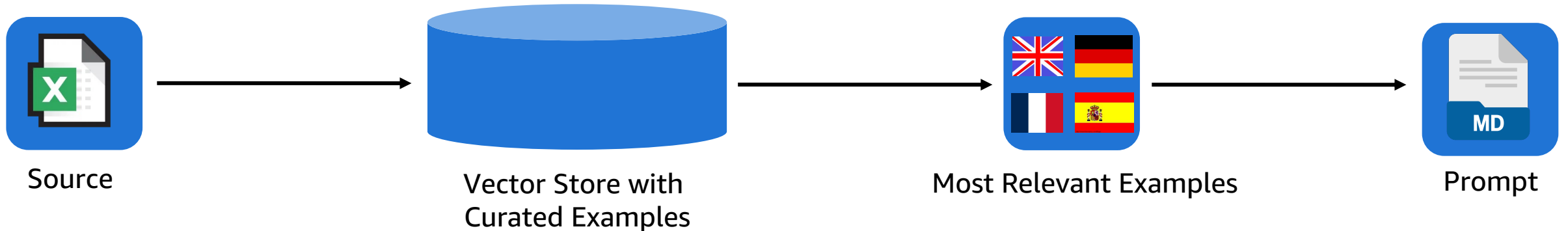
The prompt begins with a detailed persona ("You are a highly skilled copywriter at Zalando...") to set the context and tone

Prioritized Instructions

The full tonality guide is placed at the start of the prompt, before other instructions, to ensure the model gives it the highest priority

Dynamic Few-Shot Learning

We don't use static examples. Instead, we perform a real-time similarity search against a **FAISS vector index** to retrieve and inject the most relevant examples for the specific content being translated



Ingredient 2: Hybrid Approach to Measuring Quality

Stage 1: Automated & Scalable with LLM-as-a-Judge

Process: We used an LLM to evaluate the generated content, comparing it to both the ground truth and the tonality guide.

Key Finding: The LLM-as-a-Judge was significantly more useful than standard similarity metrics like BLEU. **It understands stylistic and tonal adherence**, whereas similarity metrics only measure word overlap and often penalize good, creative localizations.

Stage 2: Human-in-the-Loop Feedback

Process: Human subject matter experts reviewed the AI output against 5 core criteria: accuracy, spelling, flow, formatting, and grammar.

Actionable Loop: The specific, field-level feedback from this stage was crucial for creating a direct feedback loop to iteratively improve our prompt templates



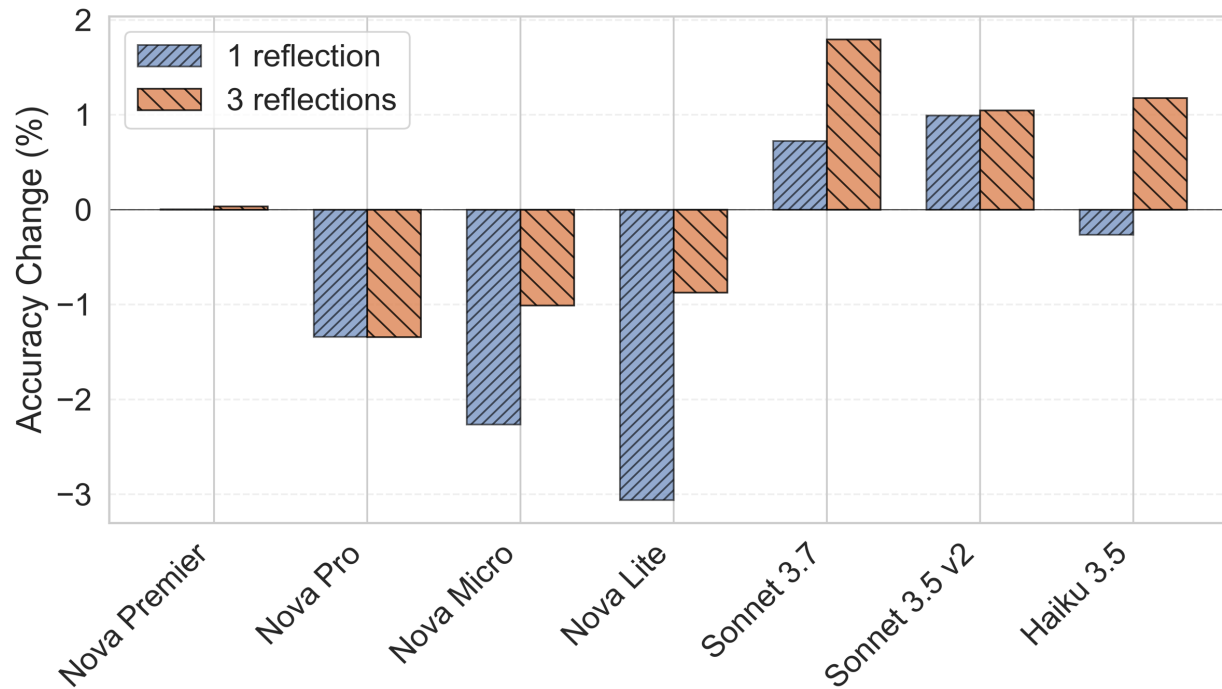
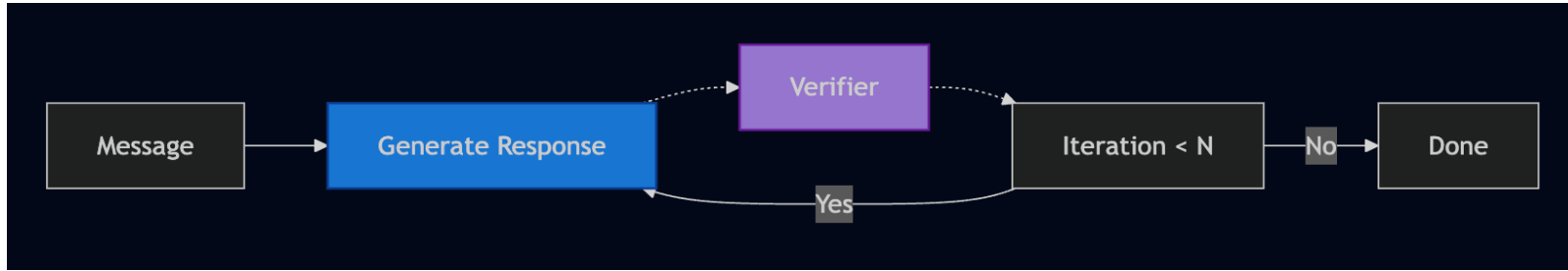
LLM as a Judge

+



Human Review

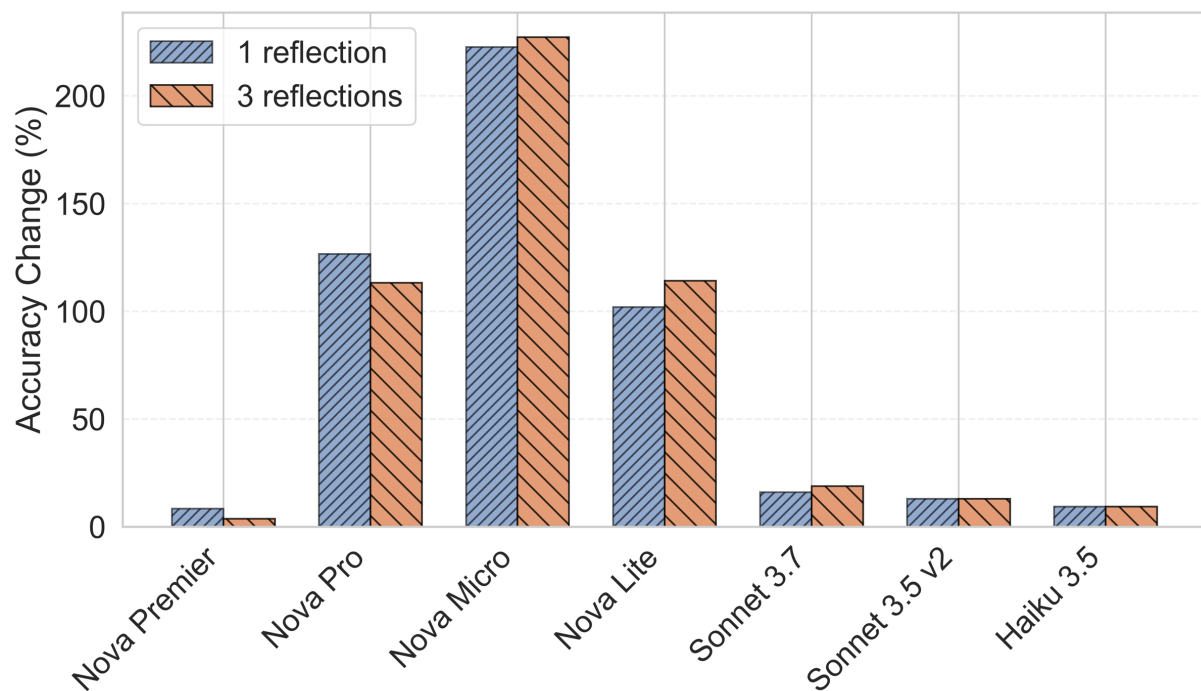
Ingredient 3: Self-Reflection



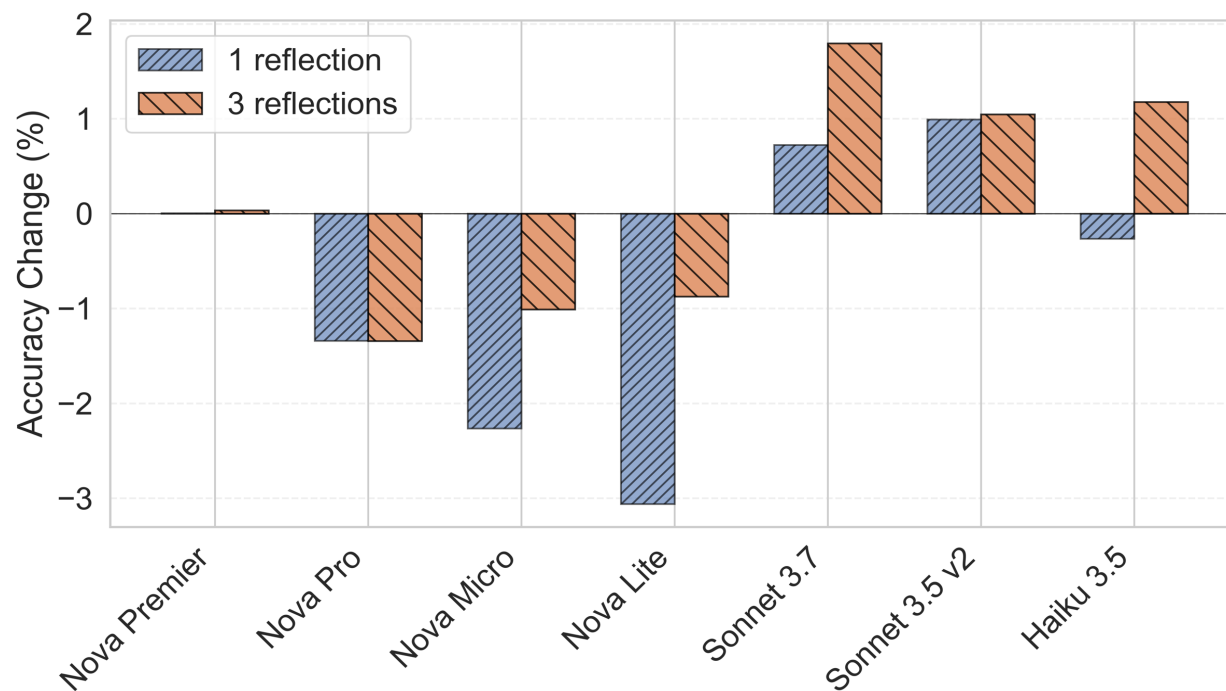
Process: LLM iteratively regenerates output after examining the feedback (generated evaluation report)

Results: Different LLMs show different performance with iterations. On the localization task, most Amazon Nova models perform well out of the box and do not benefit from extra iterations, whereas Anthropic Claude models benefit from one or more self-reflection rounds.

Selected Experiment Results



Math problems



Localization tasks

Benchmarking: conducted large-scale analysis on multiple datasets to understand trade-offs of self-reflection

Selected Experiment Results

Language	No reflection			Self-reflection with LLM judge feedback		
	BLEU	METEOR	LLM judge score	BLEU	METEOR	LLM judge score
French	0.16	0.47	0.61	0.14	0.42	0.62
Spanish	0.29	0.61	0.49	0.29	0.59	0.50
German	0.32	0.61	0.38	0.33	0.62	0.47

LLM judge: self-reflection improves the localisation quality on the selected 3 languages

Research Paper at KDD & NeurIPS Workshops

Finding the Sweet Spot: Trading Quality, Cost, and Speed During Inference-Time LLM Reflection

Jack Butler

Amazon Web Services
jackbtlr@amazon.co.uk

Nikita Kozodoi

Amazon Web Services
kozodoi@amazon.de

Zainab Afolabi

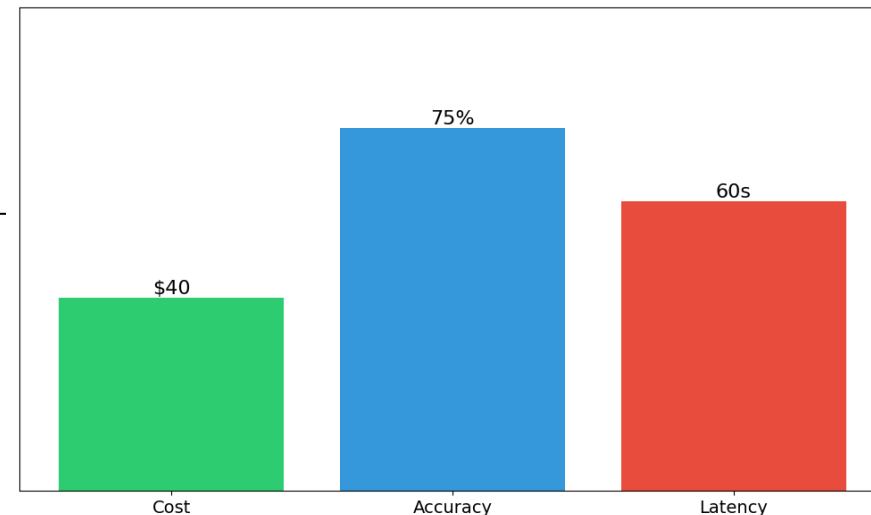
Amazon Web Services
zafolabi@amazon.co.uk

Brian Tyacke

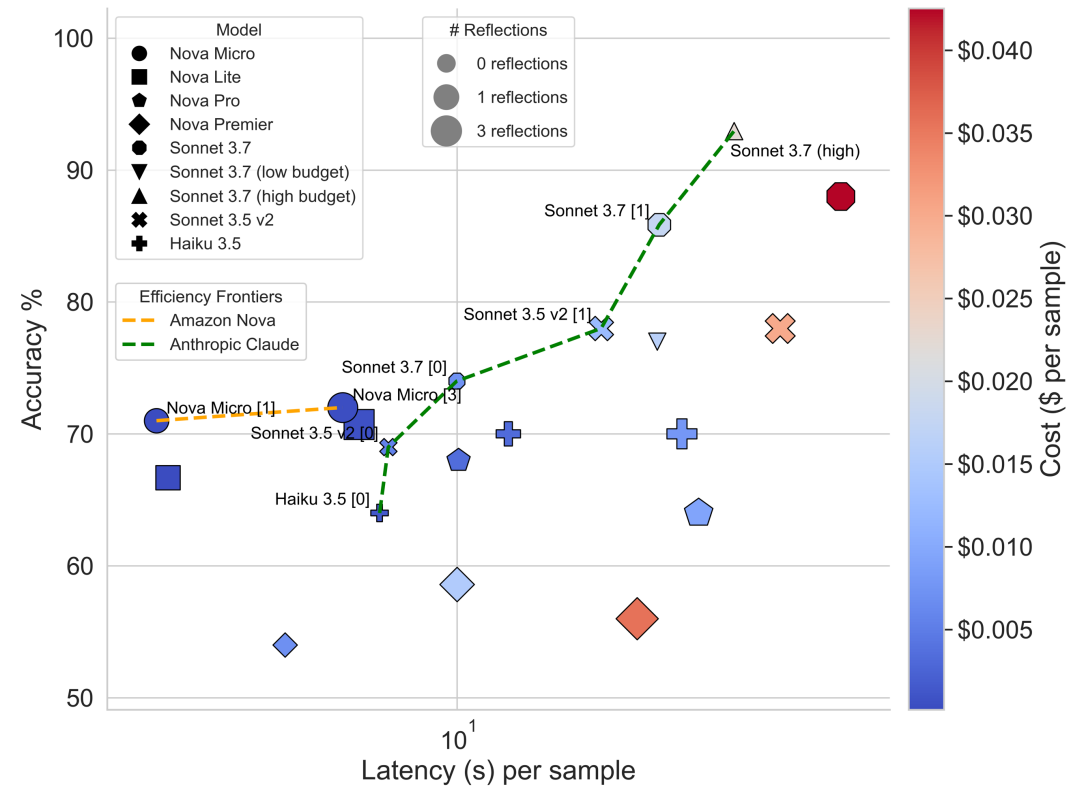
Zalando
brian.tyacke@zalando.de

Gaiar Baimuratov

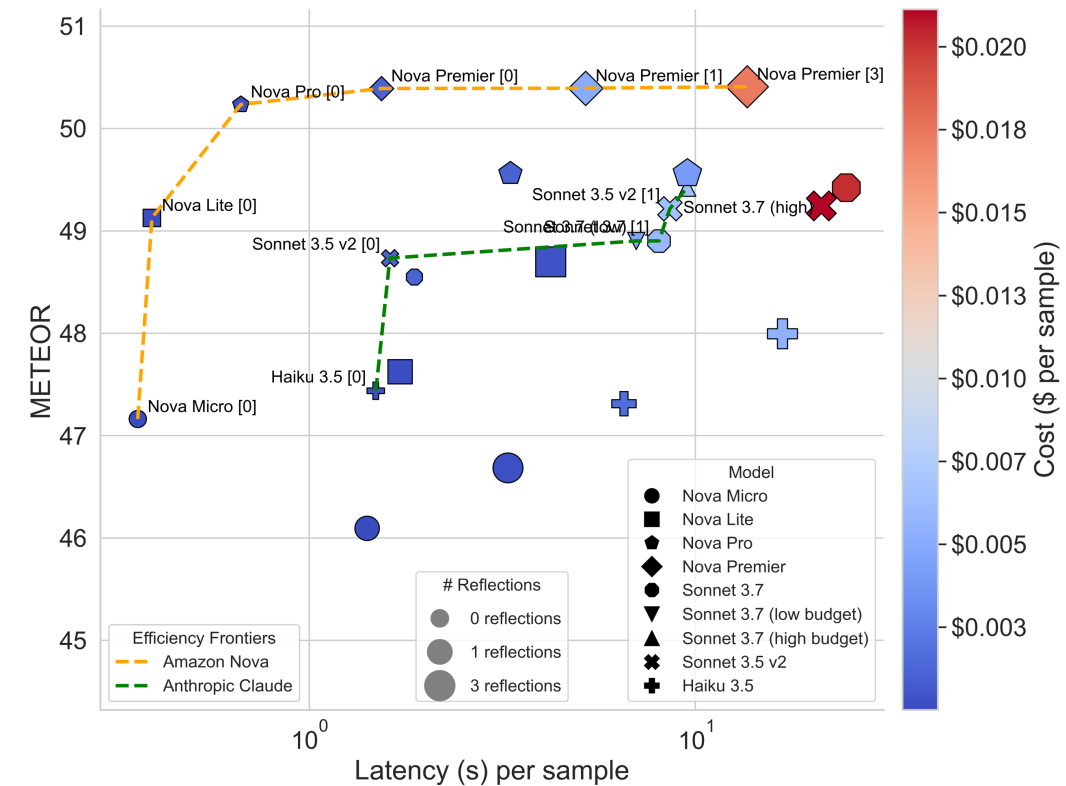
Zalando
gaiar.baimuratov@zalando.de



Selected Experiment Results



Math problems



Localization tasks

Benchmarking: conducted large-scale analysis on multiple datasets to understand trade-offs of self-reflection

CRM

92% Cost Reduction

Reducing reliance on external partners.

90% Efficiency Uplift

Through AI-driven localization processes.

Agile Content at Scale

Enables rapid, localized content for dynamic campaigns



PDPs

30% Cost Reduction

Reducing reliance on external partners.

35% Efficiency Uplift

Through AI-driven localization processes.

14k Descriptions Weekly

AI generates weekly 14,000 product descriptions on average, enhancing scalability.

Key Technical Learnings & Takeaways

On Prompting

Input Quality is Paramount: The quality and lack of conflicts in the tonality guides and few-shot examples have the single biggest impact on output quality.

Structure is Everything: The order of instructions matters. Placing the tonality guide before other instructions demonstrably improved adherence.

On Model Behavior

Controlling Creativity: A low but non-zero temperature (e.g., 0.2) was the optimal balance to avoid overly literal translations while maintaining factual accuracy and brand voice.

Self-Reflection is Powerful but Domain-Specific: Our research showed that self-revision can yield huge gains (up to 220% in some tasks) but must be validated for each use case.

On Cost

The Cost of Quality: Advanced techniques like self-reflection add cost and latency. Our research confirms that prompt caching can be an effective mitigation, offsetting the additional cost by up to 28%

Next Steps

Continuous Model Evaluation

Actively explore and benchmark the latest foundation models to ensure the solution stays current and leverages the best-performing capabilities available.

Automated Evaluation Pipelines

Develop automated evaluation pipelines to provide immediate, data-driven feedback. This is crucial for understanding the impact of every change to a prompt or tonality guide, ensuring all updates lead to measurable improvements.

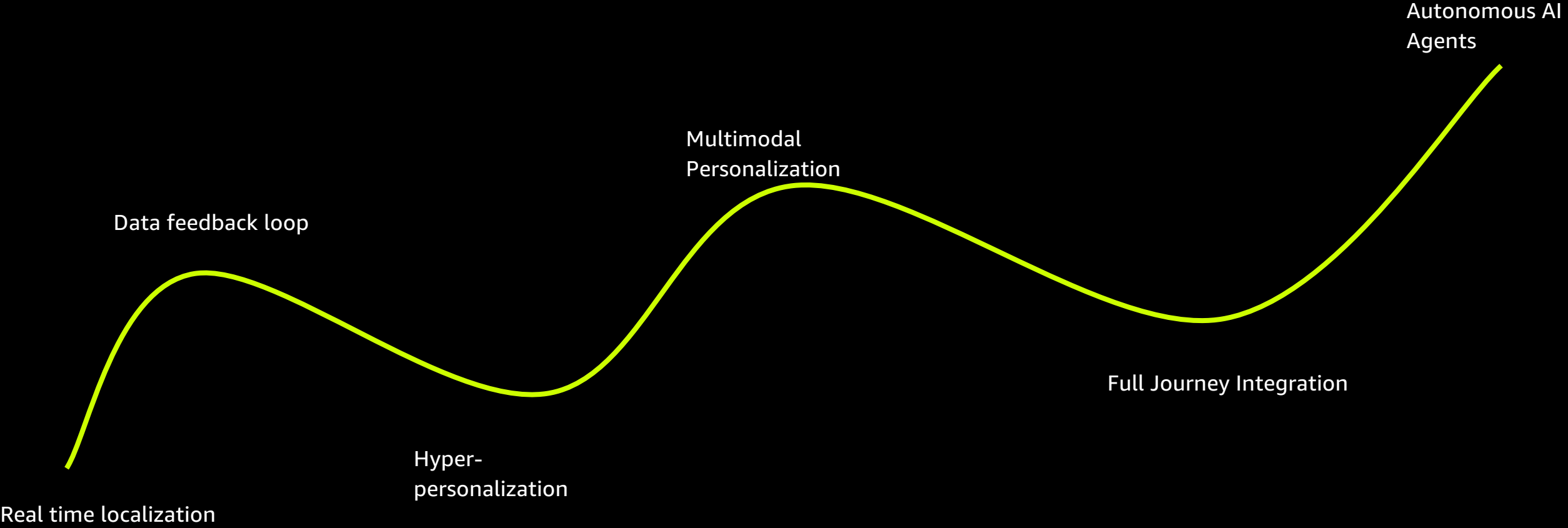
Robust Prompt Management

Integrate the solution with specialized prompt versioning and management services, such as Langfuse, to systematically track, test, and manage the lifecycle of all prompts.

Seamless CRM Integration

Embed the localization service directly into the CRM communication tech stack. This will create a seamless, end-to-end workflow, eliminating the need for manual copy-pasting of translated content and accelerating the time-to-market for marketing campaigns.

Innovating Tomorrow: GenAI in Content Creation





Thank you!

 [linkedin.com/in/gaiar](https://www.linkedin.com/in/gaiar)

Gaiar Baimuratov

Head of Product and Tech
Lounge by Zalando

 [linkedin.com/in/kozodoi](https://www.linkedin.com/in/kozodoi)

Nikita Kozodoi, PhD

Senior Applied Scientist
Amazon Web Services



Please complete the
session survey.

