



SES107

6 Things we wished we knew before

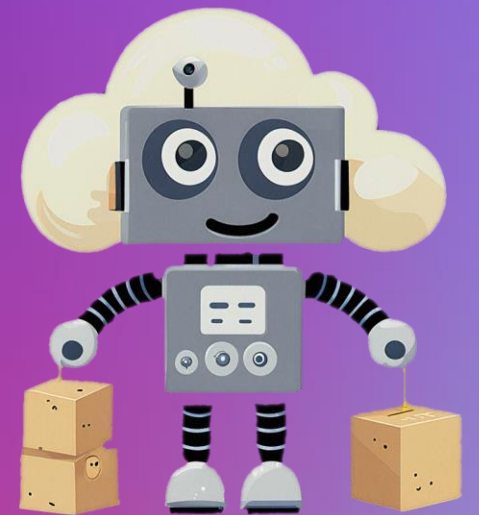
Building our first Generative AI Application

Tobias Matern

Solutions Architect
AWS

Christian Bonzelet

Solutions Architect
DFL Digital Sports



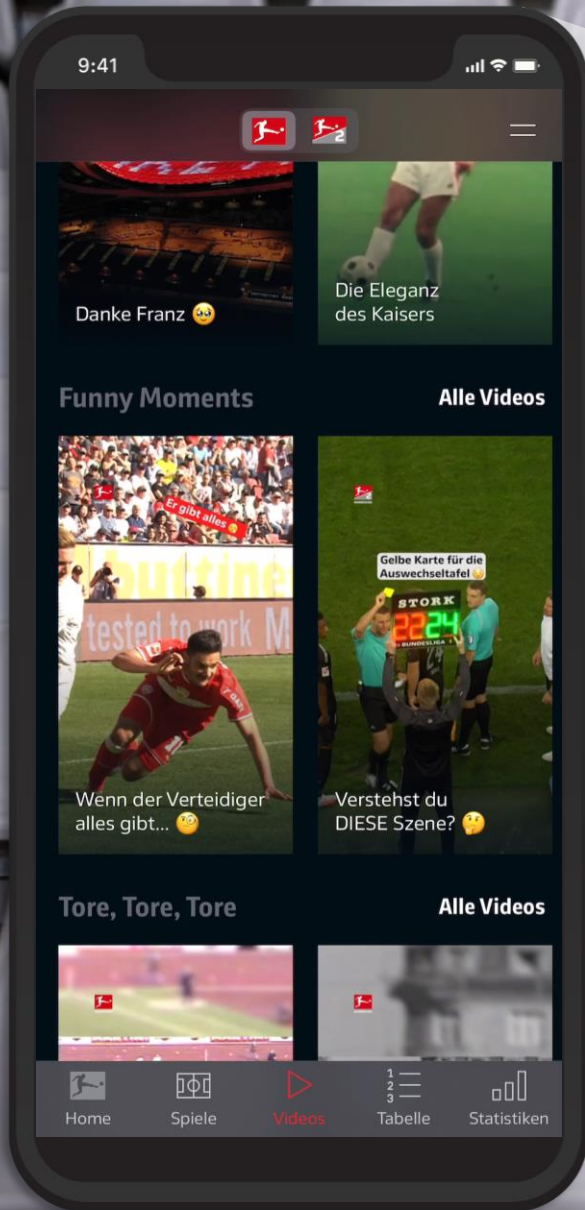
**“ THIS PROJECT WOULD NOT HAVE
BEEN POSSIBLE WITHOUT
GENERATIVE-AI ”**

Dirk Switalla

Senior Product Owner for Content Management

DFL Digital Sports

FOOTBALL IS MORE THEN A SPORT



6 THINGS WE WISHED WE KNEW BEFORE

BUNDESLIGA SHORTS

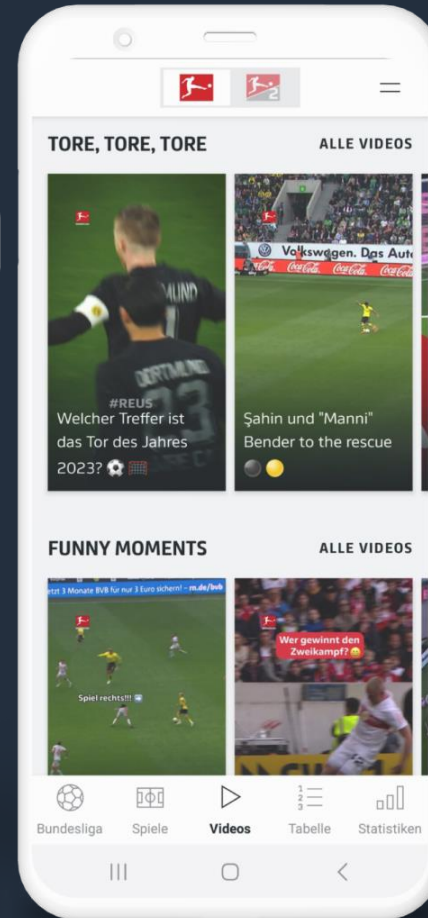
AN ENTERTAINING AND PERSONALIZED CONTENT FORMAT

20

VIDEO-PLAYS PER SESSION

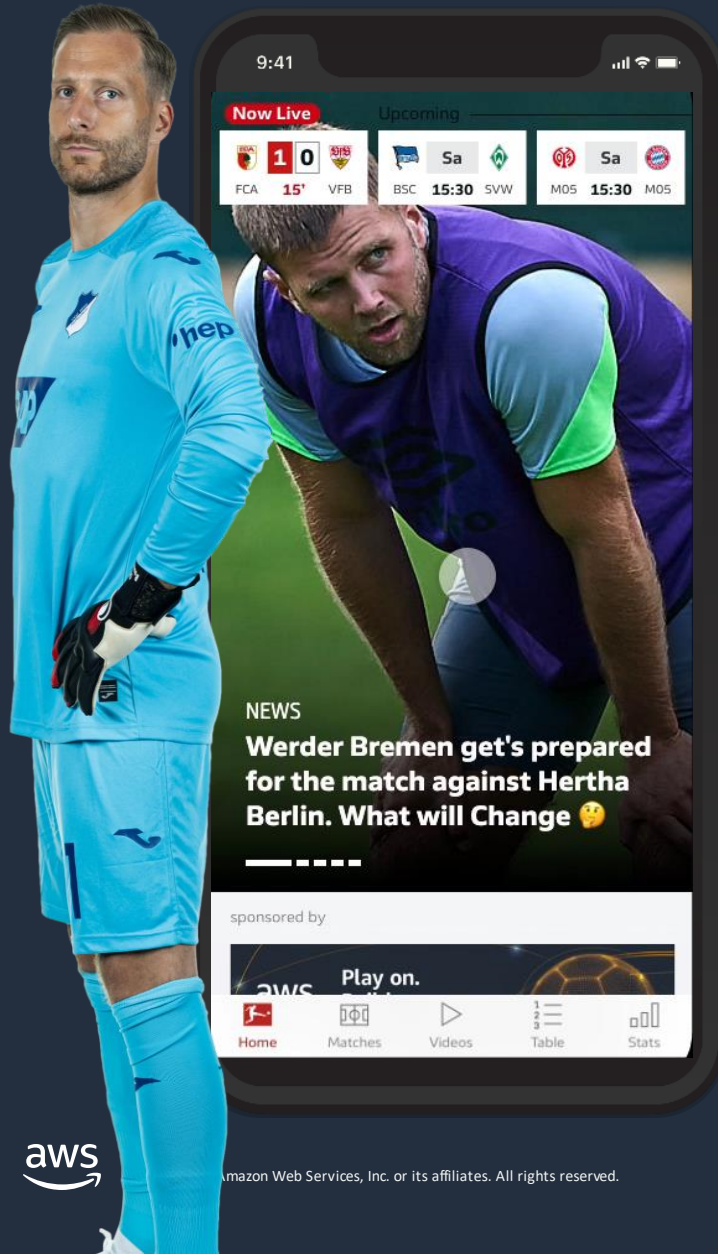
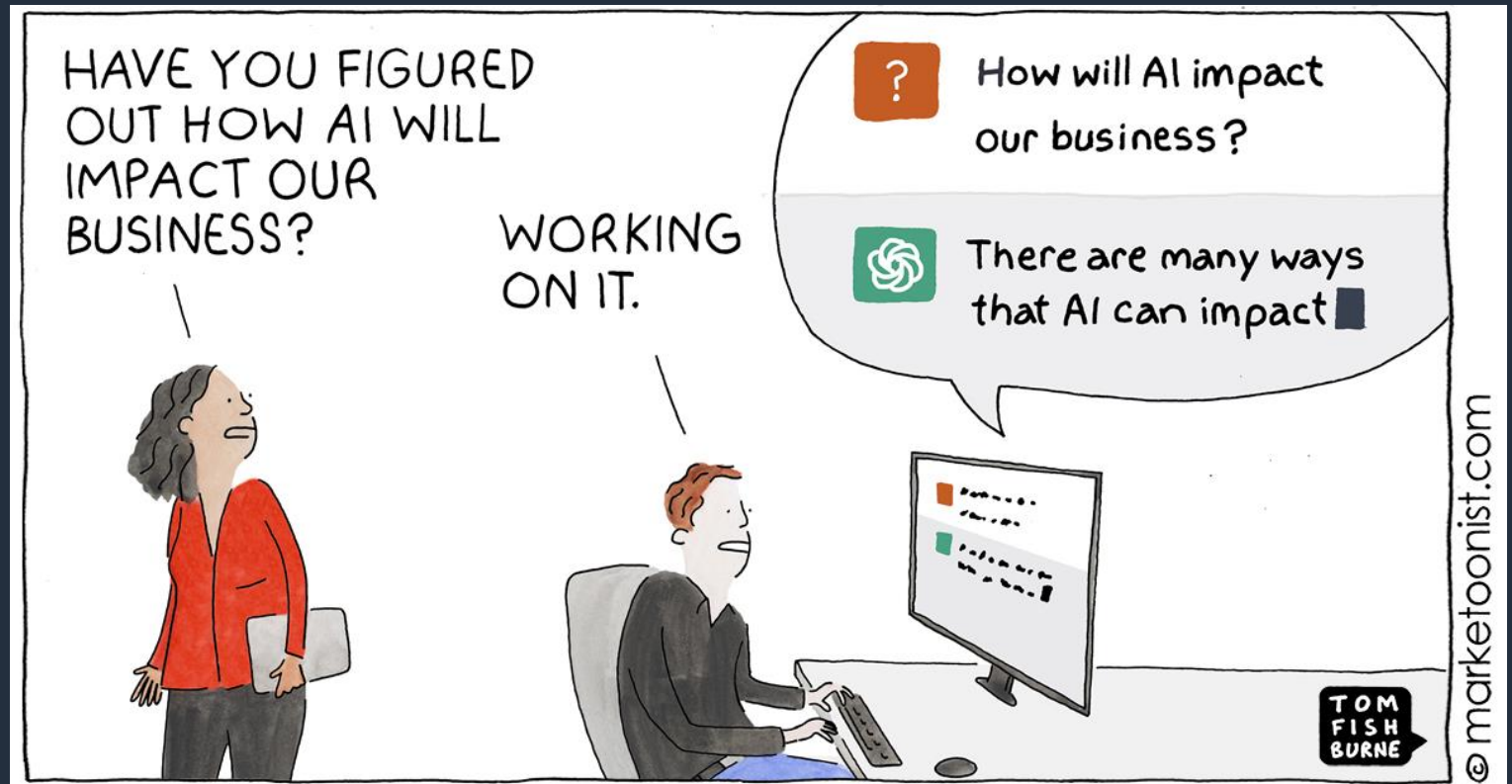
+60%

INCREASE OF TIME-SPENT IN APP



BUNDESLIGA STORIES

PROVIDES THE BEST OF BOTH WORLDS



**“ THIS PROJECT WOULD NOT HAVE
BEEN POSSIBLE WITHOUT
GENERATIVE-AI ”**

Dirk Switalla

Senior Product Owner for Content Management

DFL Digital Sports

OUR SIMPLE PLAN

LET THE AI DO OUR JOB AND US GOING STRAIGHT INTO PRODUCTION

Write a Prompt.

We want you to summarize an article and transform it into a list of discrete slides. And here is the article you should summarize:

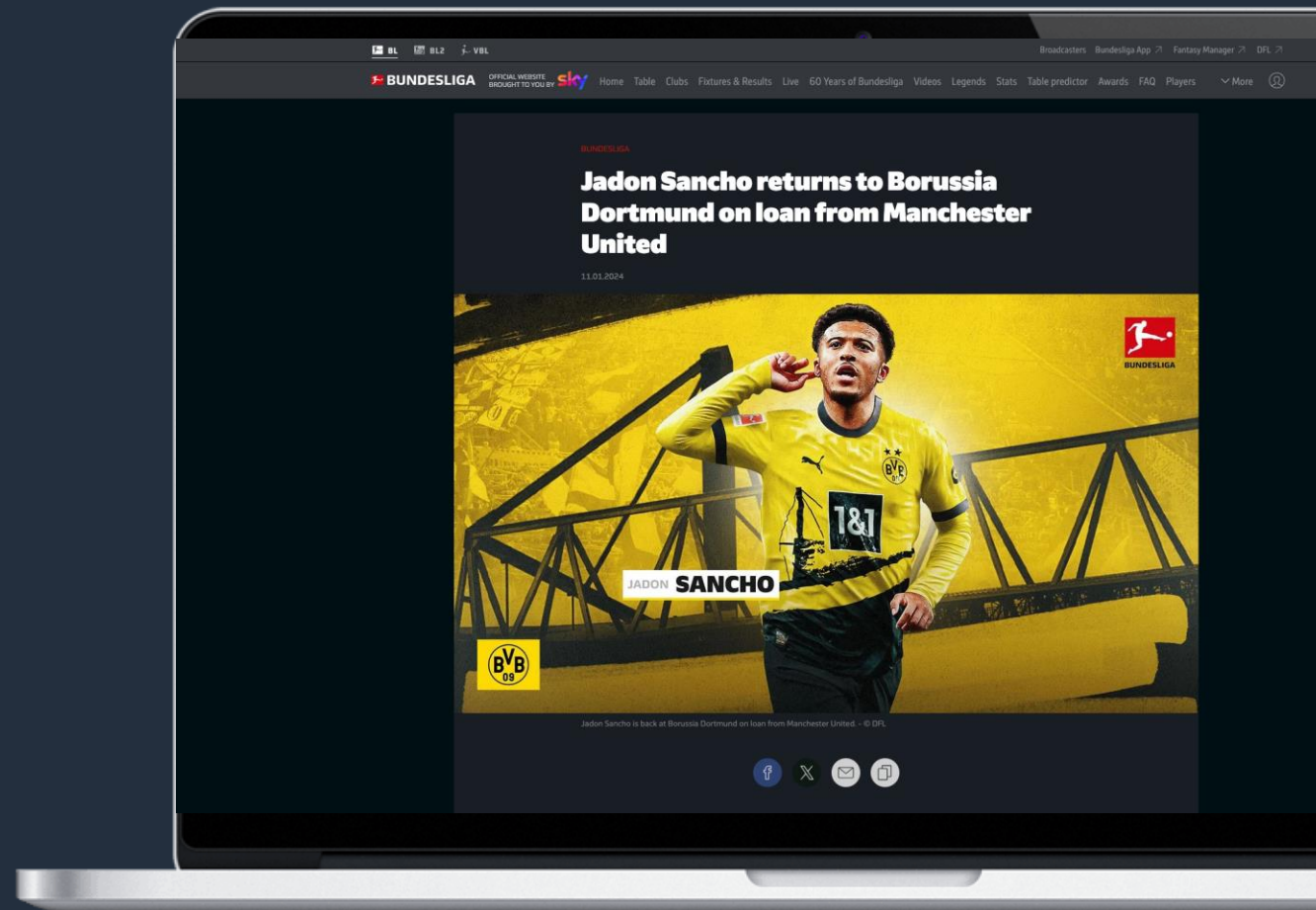
...

Launch a Party.

OUR FIRST OBSTACLE

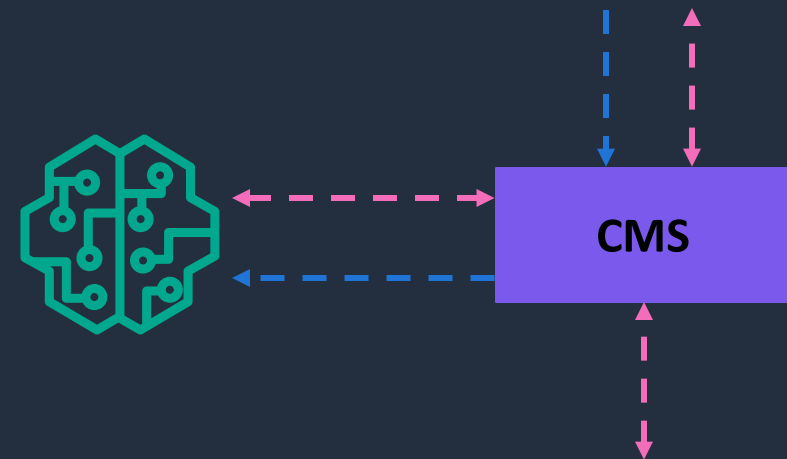
BE PRECISE IN DEFINING THE TASK THE AI SHOULD TAKE CARE OF

- 1. EXTRACTING AND SANITIZING THE TEXT OF A WEB ARTICLE FROM BUNDESLIGA.COM.
- 2. TAKE A LONG BUNCH OF TEXT AND IDENTIFY KEY ASPECTS OUT OF IT.
- 3. REWRITE KEY ASPECTS INTO A "SLIDE" CONTAINING A TITLE AND A DESCRIPTION.
- 4. FIND AN IMAGE THAT MATCH TO THE CONTEXT OF THE SLIDE AND ARTICLE.
- 5. ENSURE COMPLIANCE AND QUALITY.



THE LARGE LANGUAGE MODEL IS NOT THE APP

IT IS A USUAL COMPONENT IN A BROADER SYSTEM LANDSCAPE



--- CONTROL FLOW

--- DATA FLOW

HOW DOES AN EDITOR INTERACT WITH THE LLM?

WHAT DATA DO WE NEED TO LET THE AI DO THE JOB?

WHAT IS THE DATA FORMAT THAT WE EXCHANGE?

1. YOU ARE NOT BULDING THE NEXT BEST CHAT-BOT

WE FELT LIKE KIDS IN A TOY STORE

CHOOSING THE RIGHT FOUNDATIONAL MODEL IS THE FIRST IMPORTANT MILESTONE

AI21 labs

ANTHROPIC

cohere

∞ Meta

stability.ai

amazon

Jurassic

Contextual answers,
summarization,
paraphrasing

Claude

Summarization, complex
reasoning, writing, coding

Command + Embed

Text generation, search,
classification

Llama 2

Q&A and reading
comprehension

Stable Diffusion

High-quality images and art

Amazon Titan

Text summarization, generation,
Q&A, search

CHOOSING THE RIGHT FOUNDATIONAL MODEL

USE MODEL CARDS

Model Card and Evaluations for Claude Models

Anthropic

1 Introduction

This report includes the model card [1] for Claude models, focusing on Claude 2, along with the results of a range of safety, alignment, and capabilities evaluations. We have been iterating on the training and evaluation of Claude-type models since our first work on Reinforcement Learning from Human Feedback (RLHF) [2]; the newest Claude 2 model represents a continuous evolution from those early and less capable 'helpful and harmless' language assistants.

This report is not intended to be a scientific paper since most aspects of training and evaluating these models have been documented in our research papers. These include papers on preference modeling [3], reinforcement learning from human feedback for helpful and harmless models [2], red teaming language models [4], measuring representation of subjective global values in language models [5], honesty, (i.e., exploring language models' ability to recognize what they know) [6], evaluating language models with language model-generated tests [7], moral self-correction [8], and Constitutional AI [9]. We also discussed Claude's specific constitution in a recent blog post [10]. Our work using human evaluations to test model safety is most thoroughly documented in our paper "Red-Teaming Language Models to Reduce Harms" [4], while our recent work on automated safety evaluation is "Discovering Language Model Behaviors with Model-Written Evaluations" [7].

This report is also not comprehensive – we expect to release new findings as we continue our research and evaluations of frontier models. However, we hope it provides useful insight into Claude 2's capabilities and limitations.

Update (November 21, 2023): We have added an appendix section covering Claude 2.1.

2 Claude 2 Model Card

Claude 2 is our most capable system yet, and we hope it will unlock a range of new and valuable use cases. That said, the model is far from perfect. In this model card, we hope to display Claude 2's strengths and limitations as well as describe the evaluations and safety interventions we have conducted to improve helpfulness, honesty, and harmlessness (HHH).

THE MODELS STRENGTH ALIGN WITH OUR REQUIREMENTS.

ETHICAL CONSIDERATIONS OF THE MODEL ALIGN WITH OUR CORPORATE VALUES.

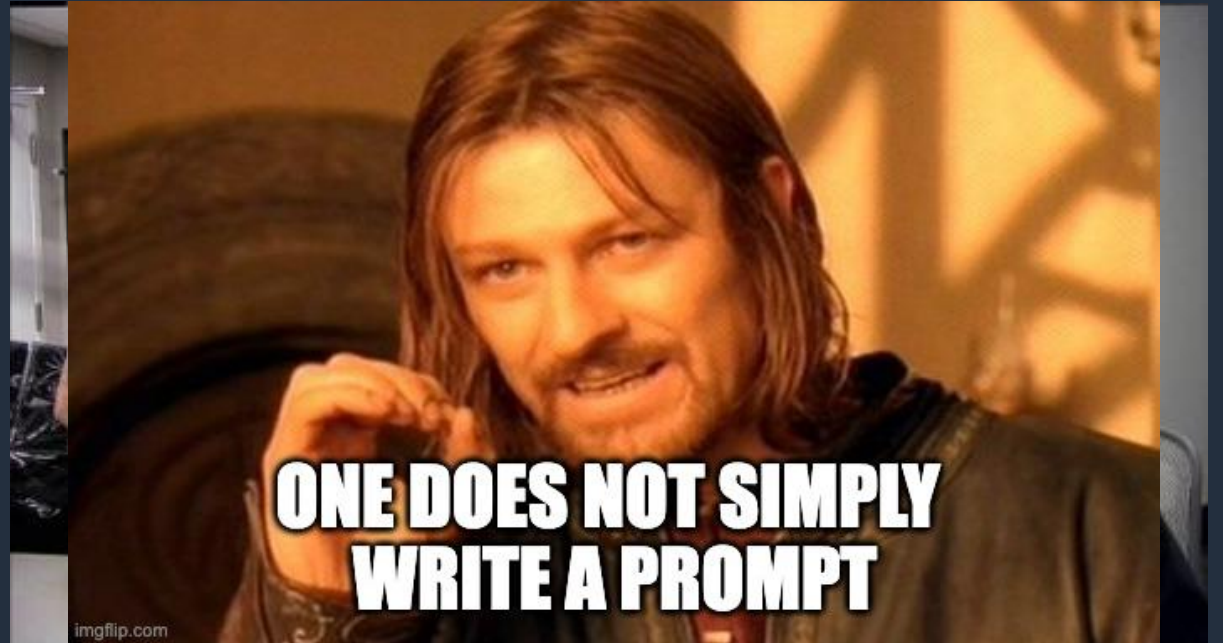
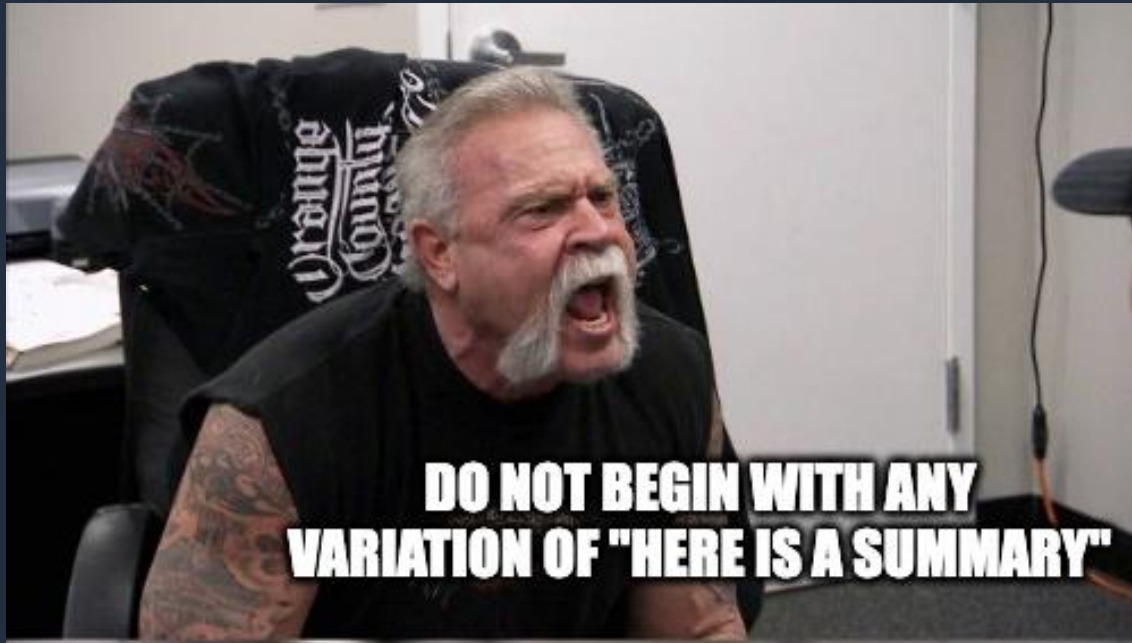
MODELS STEERABILITY ENSURES FLEXIBILITY.



PROVIDED VIA
AMAZON BEDROCK

CONSIDER YOUR PROMPT AS AN API

DEAR CLAUDE, I WANT ONLY JSON.



THE SYMBIOSIS OF A PROMPT AND ITS MODEL

THE PROMPT YOU WRITE FOR MODEL A MIGHT NOT WORK AS EXPECTED FOR MODEL B

WRITE. TEST. REFINE. REPEAT.

DESCRIBE YOUR TASK WELL.

MARK DIFFERENT PARTS OF
YOUR PROMPT WITH XML.

GIVE SOME EXAMPLES.

LET CLAUDE “THINK”.



```
1 <thinking>
2 Here is my plan for summarizing the article into slides:
3 - Extract the key points from each paragraph
4 - Come up with a title for each key point
5 - Write a short description for each key point
6 </thinking>
7
8 <answer>
9 [
10   {
11     "title": "Flawless Performance from Schalke",
12     "description": "Schalke put on a clinical display, scoring f
13   },
14   {
15     "title": "Memorable Debut Goal for Matondo",
16     "description": "18-year-old Rabbi Matondo scored his first e
17   },
18   {
19     "title": "Growing Confidence for Schalke",
```

2. THE PROMPT IS NOT ENOUGH

HOW DOES IMAGE SELECTION PERFORM?

IT IS OKAY. BUT DFL WANTS TO PUSH FOR MORE



Schalke's Flawless Performance Against Leipzig

Schalke were clinical in their 3-0 win over league leaders Leipzig. They scored from a corner, penalty and counter-attack.



Matondo's memorable first Bundesliga Goal

18-year-old Rabbi Matondo scored his first Bundesliga goal in his 8th league appearance for Schalke.



Confidence high ahead

With wins over Cologne, Hoffenheim and Dortmund coming up, Matondo is confident Schalke's purple patch can continue.

DATA IS YOUR DIFFERENTIATOR



Enjoy Generative AI
models with
NO CUSTOM DATA



Extend model
capabilities using
**CUSTOM DATA AT
HAND**



Enlight model
capabilities using
**HIGH-QUALITY
CUSTOM DATA AT
HAND**

THESE ARE FOOTBALL PLAYERS. NOT DANCERS

WHAT WE KNOW ABOUT OUR IMAGES IS WHAT AN AI KNOWS ABOUT OUR IMAGES



COPYRIGHT:

imago images / Nordphotopyright

CAPTION:

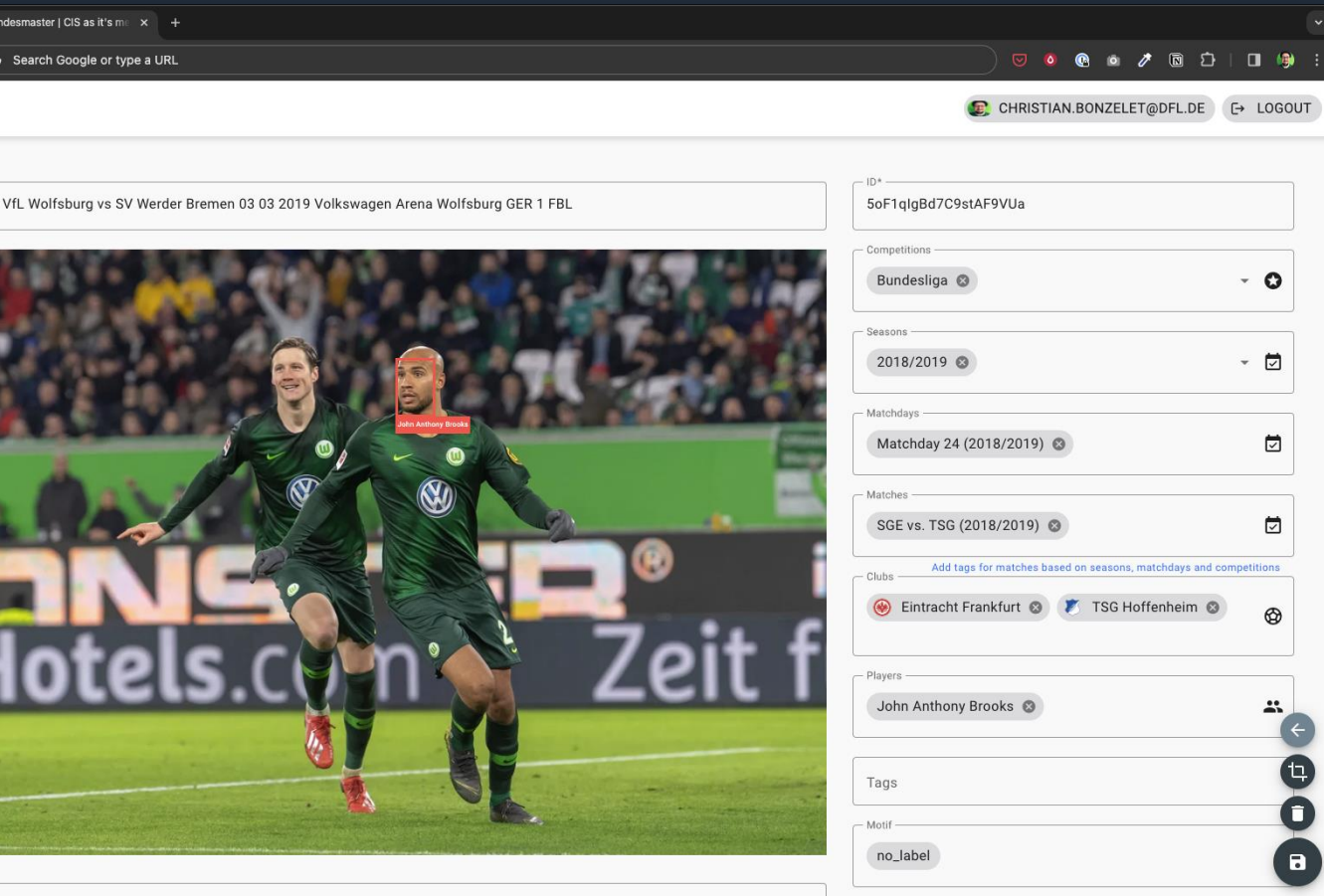
John Brooks (r.) scored his second goal of the season as Wolfsburg shared the spoils with Werder Bremen on Sunday.

ALTERNATIVE TEXT:

John Brooks

IMPROVED METADATA OF IMAGES

WITH THE HELP OF AMAZON REKOGNITION



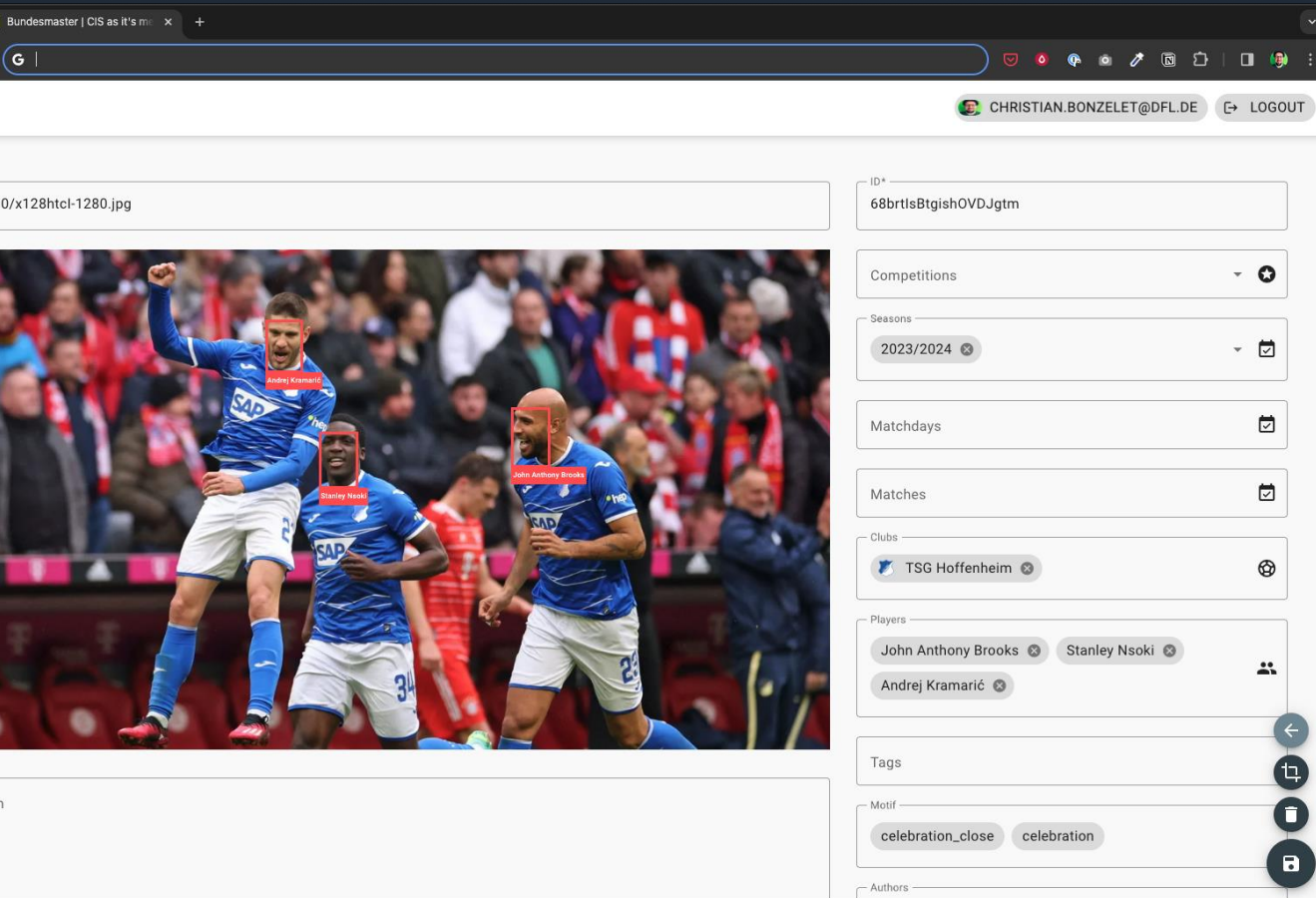
116.000
analyzed images

954.000
detected faces

200.000
detected players / staff members

IMPROVED METADATA OF IMAGES

WITH THE HELP OF AMAZON REKOGNITION



116.000

processed images

15

custom image classifications

132.000

detected custom labels

3. SHIT IN. SHIT OUT.



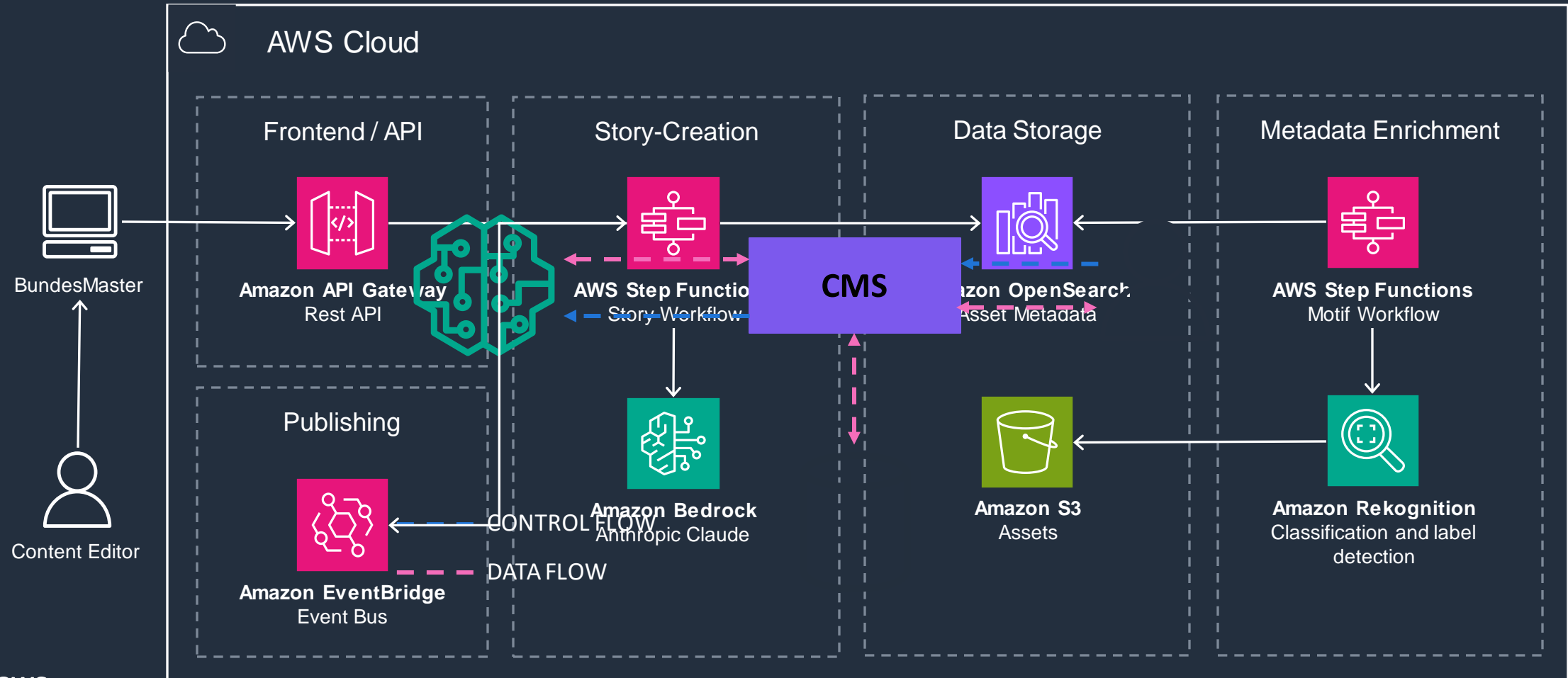
AUTOMATED 9:16 CROPPING

GENAI IS NOT THE ANSWER TO ALL QUESTIONS



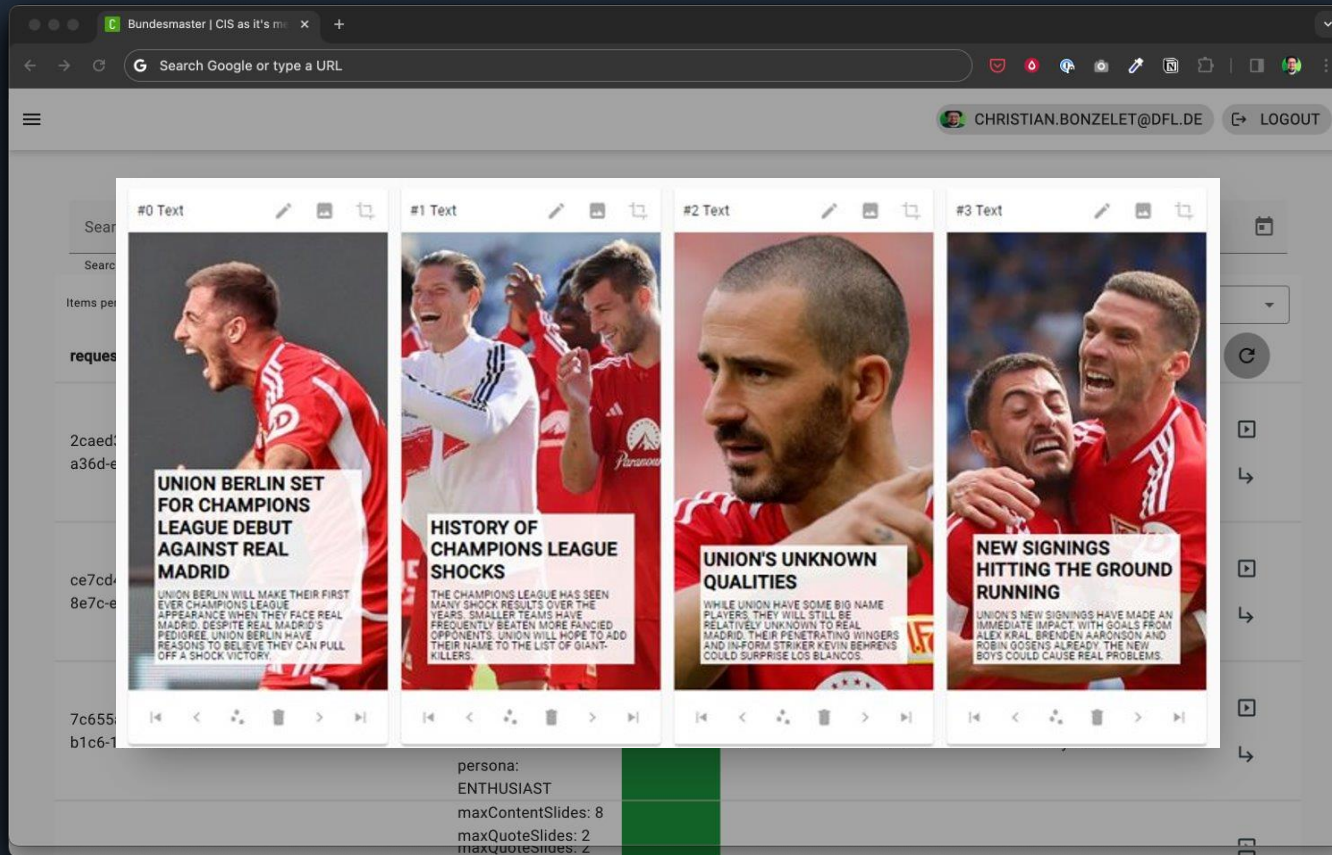
OUR SOLUTION ARCHITECTURE

A CUSTOM-BUILT CMS WITH AN INTEGRATED LLM



THE CREATION PROCESS IN A NUTSHELL

AI-GENERATED AND HUMAN-REFINED BUNDESLIGA STORIES



SELECT AN ARTICLE THAT SHOULD BE TRANSFORMED INTO A STORY FORMAT.

START THE STORY GENERATION JOB AND GRAB A COFFEE.

VERIFY THE STORY SUGGESTION THE AI GENERATED FOR YOU.

MAKE CHANGES, PUBLISH AND ENGAGE OUR FANS.

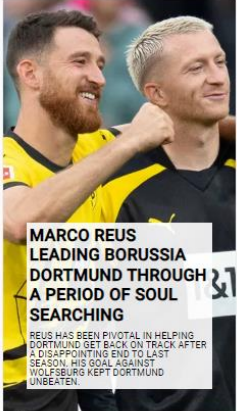

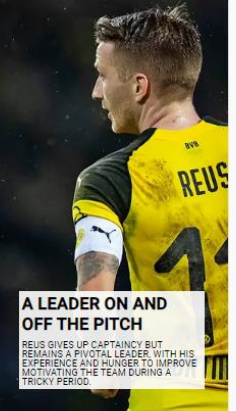

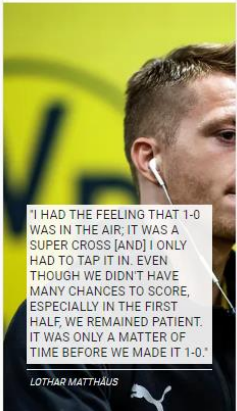
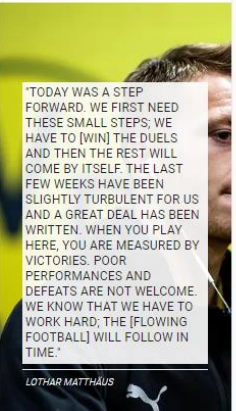


4. IT IS NOT THE LLM TO RULE THEM ALL

MODEL EVALUATION

HOW DO WE KNOW OUR MODEL PERFORMS WELL?

Slide #2 Verschiedene Bacon Arten no image Der Artikel listet viele verschiedene Bacon Sorten auf, wie z.B. landjaeger, chicken, buffalo, tongue, etc. Type: text Actors: Autor Scene: stadium Image Source: article	Slide #3 Noch mehr Bacon no image Im zweiten Abschnitt werden weitere Bacon Varianten aufgezählt, wie brisket, pork shank, salami, andouille, etc. Type: text Actors: Autor Scene: stadium Image Source: article	Slide #4 Bacon Rezepte no image Verschiedene Rezepte mit Bacon als Hauptzutat werden vorgestellt, z.B. hamburger, corned beef, sausage, etc. Type: text Actors: Autor Scene: player Image Source: article
Slide #6 Bacon Verarbeitung no image Der Artikel erklärt, wie Bacon verarbeitet und zubereitet wird, z.B. als Sausage, Salami, Burgdoggen, etc. Type: text Actors: Autor Scene: action Image Source: article	Slide #7 Noch mehr Verarbeitung no image Weitere Details zur Bacon Verarbeitung werden genannt, wie die Herstellung von Pastrami, Salami, Leberkas, etc. Type: text Actors: Autor Scene: action Image Source: article	Slide #8 Fazit no image Der Artikel scheint Bacon in all seinen Varianten und Verarbeitungsformen zu feiern. Type: text Actors: Autor Scene: celebration Image Source: article

VS

#0 Text  <p>MARCO REUS LEADING BORUSSIA DORTMUND THROUGH A PERIOD OF SOUL SEARCHING REUS HAS BEEN PIVOTAL IN HELPING DORTMUND GET BACK ON TRACK AFTER A DISAPPOINTING END TO LAST SEASON. HIS GOAL AGAINST WOLFSBURG KEPT DORTMUND UNBEATEN.</p>	#1 Text  <p>REUS MOVES UP DORTMUND'S ALL-TIME SCORING LIST REUS' GOAL AGAINST WOLFSBURG WAS HIS 11TH BUNDESLIGA GOAL FOR DORTMUND, MOVING HIM INTO 3RD ON THEIR ALL-TIME SCORING LIST.</p>	#2 Text  <p>A LEADER ON AND OFF THE PITCH REUS GIVES UP CAPTAINCY BUT REMAINS A PIVOTAL LEADER. WITH HIS EXPERIENCE AND HUNGER TO IMPROVE THE TEAM AHEAD OF FACING A TRICKY PERIOD.</p>	#3 Text  <p>CALM AND ASSURED REUS' CALM WORDS AFTER HIS WOLFSBURG MATCH WILL REASSURE THE TEAM AHEAD OF FACING AN OPPONENT IN HOFFENHEIM.</p>
#4 Quote  <p>"I HAD THE FEELING THAT 1-0 WAS IN THE AIR; IT WAS A SUPER CROSS [AND] I ONLY HAD TO TAP IT IN. EVEN THOUGH WE DIDN'T HAVE MANY CHANCES TO SCORE, ESPECIALLY IN THE FIRST HALF, WE REMAINED PATIENT. IT WAS ONLY A MATTER OF TIME BEFORE WE MADE IT 1-0." LOTHAR MATTHÄUS</p>	#5 Quote  <p>"TODAY WAS A STEP FORWARD. WE FIRST NEED THESE SMALL STEPS; WE HAVE TO [WIN] THE DUELS AND THEN THE REST WILL COME BY ITSELF. THE LAST FEW WEEKS HAVE BEEN SLIGHTLY TURBULENT FOR US AND A GREAT DEAL HAS BEEN WRITTEN. WHEN YOU PLAY HERE, YOU ARE MEASURED BY VICTORIES. POOR PERFORMANCES AND DEFEATS ARE NOT WELCOME. WE KNOW THAT WE HAVE TO WORK HARD; THE [FLOWING FOOTBALL] WILL FOLLOW IN TIME." LOTHAR MATTHÄUS</p>	#6 Text  <p>A VITAL PRESENCE REUS' GOAL AGAINST WOLFSBURG DEMONSTRATES HIS CALM, ASSURED WORDS AND LEADERSHIP DURING A TRICKY PERIOD VITAL FOR DORTMUND'S REVIVAL.</p>	#7 Text  <p>LEADING DORTMUND THROUGH TURBULENT TIMES REUS' GOAL AGAINST WOLFSBURG DEMONSTRATES HIS CALM, ASSURED WORDS AND LEADERSHIP DURING A TRICKY PERIOD VITAL FOR DORTMUND'S REVIVAL.</p>

WHAT MAKES A STORY A GOOD STORY?

PRECISION AND ACCURACY:

Reference (Human):

Coutinho scored against Paderborn to go top of the Bundesliga.

Generated Output:

Coutinho provided a goal and assist as Bayern beat Paderborn 3-2 away to go top of the Bundesliga.

ROUGE-1 Precision:

0.44 (unigram matches / unigrams in output)

CONTENT EDITOR ACCEPTANCE

90%

STORY-UNTOUCHED-BY-EDITORS RATE

<2 MIN

STORY REVIEW TIME

5. DO NOT LOOSE HUMAN FEEDBACK

DATA IS YOUR DIFFERENTIATOR



Enjoy Generative AI models with
NO CUSTOM DATA



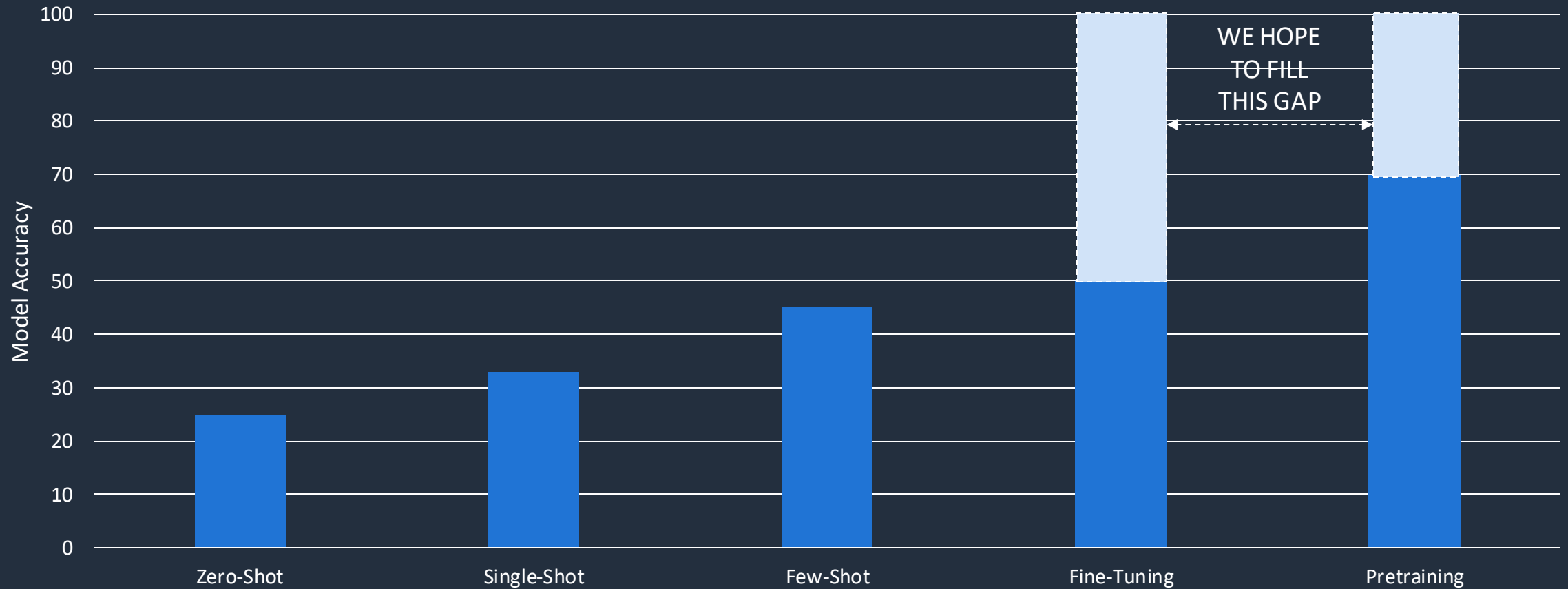
Extend model capabilities using
CUSTOM DATA AT HAND



Enlighten ~~Expand~~ the model capabilities using
HIGH-QUALITY PRETRAINING / CUSTOM DATA TUNING AT HAND

DO THOSE MODELS SPEAK FOOTBALL?

TO CONSIDER COST-INTENSIVE CUSTOMIZATIONS, YOU WANT A CHART LIKE THIS



6. KEEP ON PROMPTING

1. GENERATIVE AI IS NOT THE ANSWER TO ALL QUESTIONS
2. THE PROMPT IS NOT ENOUGH
3. SHIT IN. SHIT OUT.
4. IT IS NOT THE LLM TO RULE THEM ALL.
5. DO NOT LOOSE HUMAN FEEDBACK.
6. KEEP ON PROMPTING



Thank you!

Tobias Matern
Solutions Architect
AWS

Christian Bonzelet
Solutions Architect
DFL Digital Sports



Please complete the session survey.

