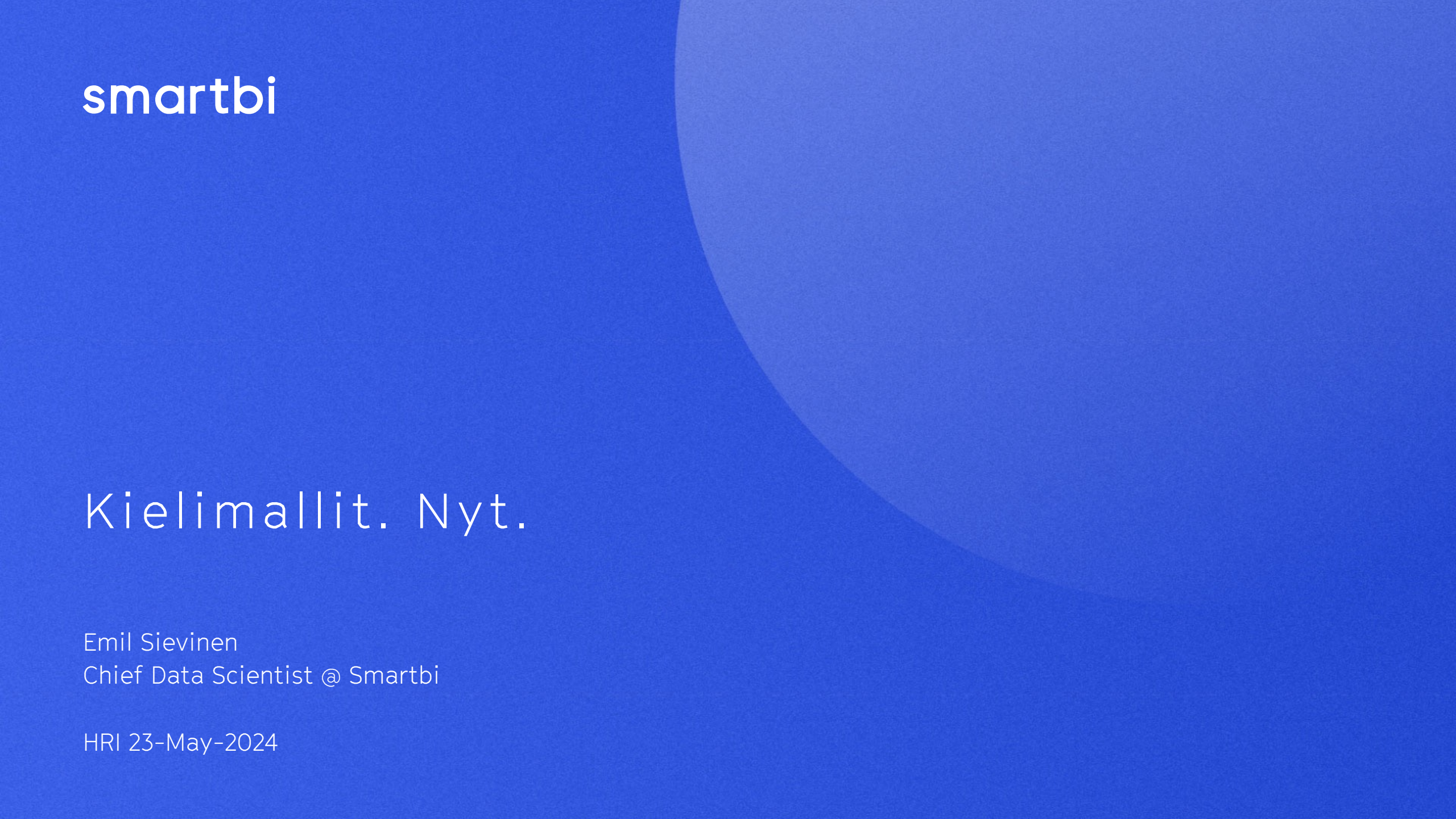




smartbi

Kielimallit. Nyt.

Emil Sievinen
Chief Data Scientist @ Smartbi

HRI 23-May-2024

Agenda

- Kielimallit nyt – yleiskatsaus markkinan megatrendeihin
- Tekniset suuntaukset kielimalleissa
- Kielimallien lähitulevaisuus
- Miten aloittaa kielimallien hyödyntäminen

Tavoitteena rakentaa parempaa
maailmaa teknologian avulla.

SMARTBI IN BRIEF

Smartbi is an advanced analytics & AI consultancy based in Finland and the United States. We design and develop tailor-made predictive and advanced process analytics, computer vision and NLP / Generative AI solutions.

Optimizing industrial
processes

Tailor-made ML products and
services

Predictive models, computer
vision, GenAI

OUR AREAS OF EXPERTISE

DESIGN

Human-centered design to solve actual business problems.

- Need and opportunity assessments.
 - Concept and business design.
 - User interface and experience design.
 - Outsourced product and project management.
-

MODELLING

Combining process knowledge with data science expertise.

- Data collection and engineering.
 - Analysis and modelling pre-study, combined with statistical and automated methods.
 - Tailored ML algorithm development.
 - Extensive model validation & quality assurance.
-

SOFTWARE

Production-grade, integrated solutions.

- End-to-end solution and integration architecture.
- Extended stack (frontend + backend + DevOps + MLOps) capability.
- Secure cloud (Azure & AWS) and EDGE ready deployments.
- Solution lifecycle management and maintenance.

SOME EXAMPLES OF OUR WORK



Strategic advanced analytics design and development partnership

Multinational pulp and
paper company

Advanced analytics program for production disturbance recovery optimization



Computer vision solution for automated preventative maintenance inspections



Computer vision solution for equipment field inspections

American agricultural
equipment manufacturer
(Fortune 100)

Predictive maintenance solution for component failure prediction and maintenance interval optimization

KIELIMALLIT NYT

LLM	Kehittäjä	Käytössä esim.	API / Avoin lähdekoodi
<u>GPT</u>	OpenAI	Microsoft, Duolingo, Stripe, Zapier, Dropbox, ChatGPT	API
<u>Gemini</u>	Google	Gemini chatbot, Google:n sovellukset	API
<u>Gemma</u>	Google		Avoin
<u>Llama 3</u>	Meta	Meta:n sovellukset, Meta AI chatbot	Avoin
<u>Vicuna</u>	LMSYS Org	Chatbot Arena	Avoin
<u>Claude 3</u>	Anthropic	Slack, Notion, Zoom	API
<u>Stable Beluga</u>	Stability AI		Avoin
<u>StableLM 2</u>	Stability AI		Avoin
<u>Coral</u>	Cohere	HyperWrite, Jasper, Notion, LongShot	API
<u>Falcon</u>	Technology Innovation Institute		Avoin
<u>DBRX</u>	Databricks and Mosaic		Avoin
<u>Mixtral 8x7B and 8x22B</u>	Mistral AI		Avoin
<u>XGen-7B</u>	Salesforce		Avoin
<u>Grok</u>	xAI	Grok Chatbot	Avoin

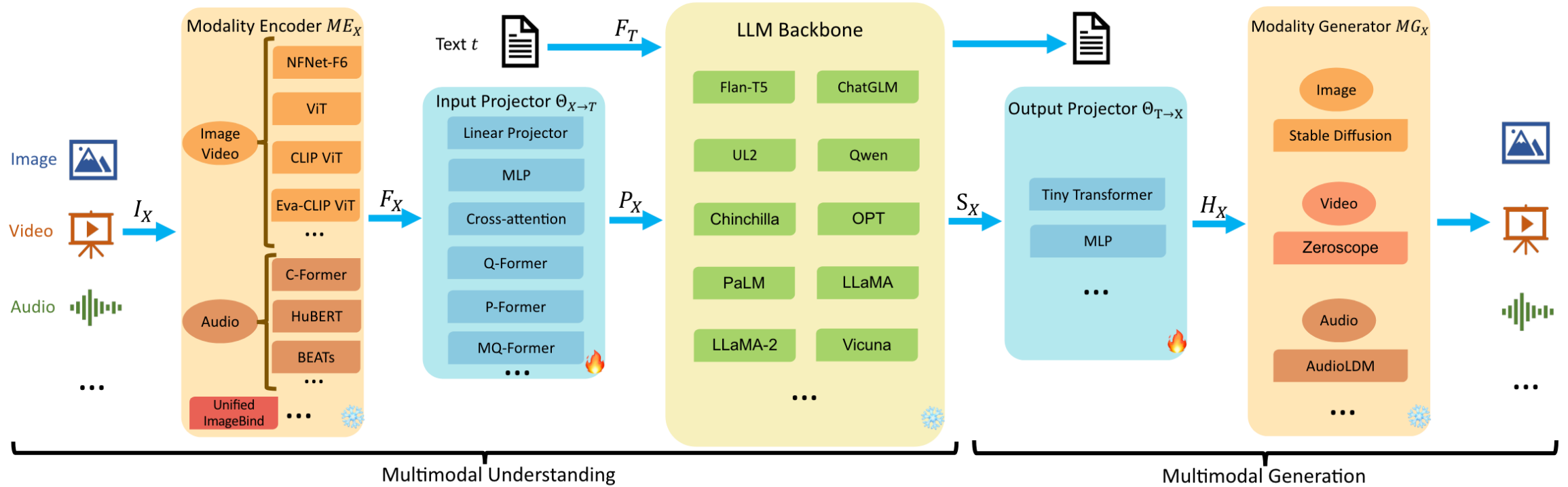
3 megatrendiä kielimallimarkkinalla

1. Avoimen lähdekoodin kielimallit tulevat tasoihin suljettujen mallien kanssa
2. Inferenssikustannukset pienenevät
3. Kielimalleja upotetaan entistä enemmän osaksi muita sovelluksia

TEKNISET SUUNTAUKSET

Keskeiset tekniset kehityssuunnat

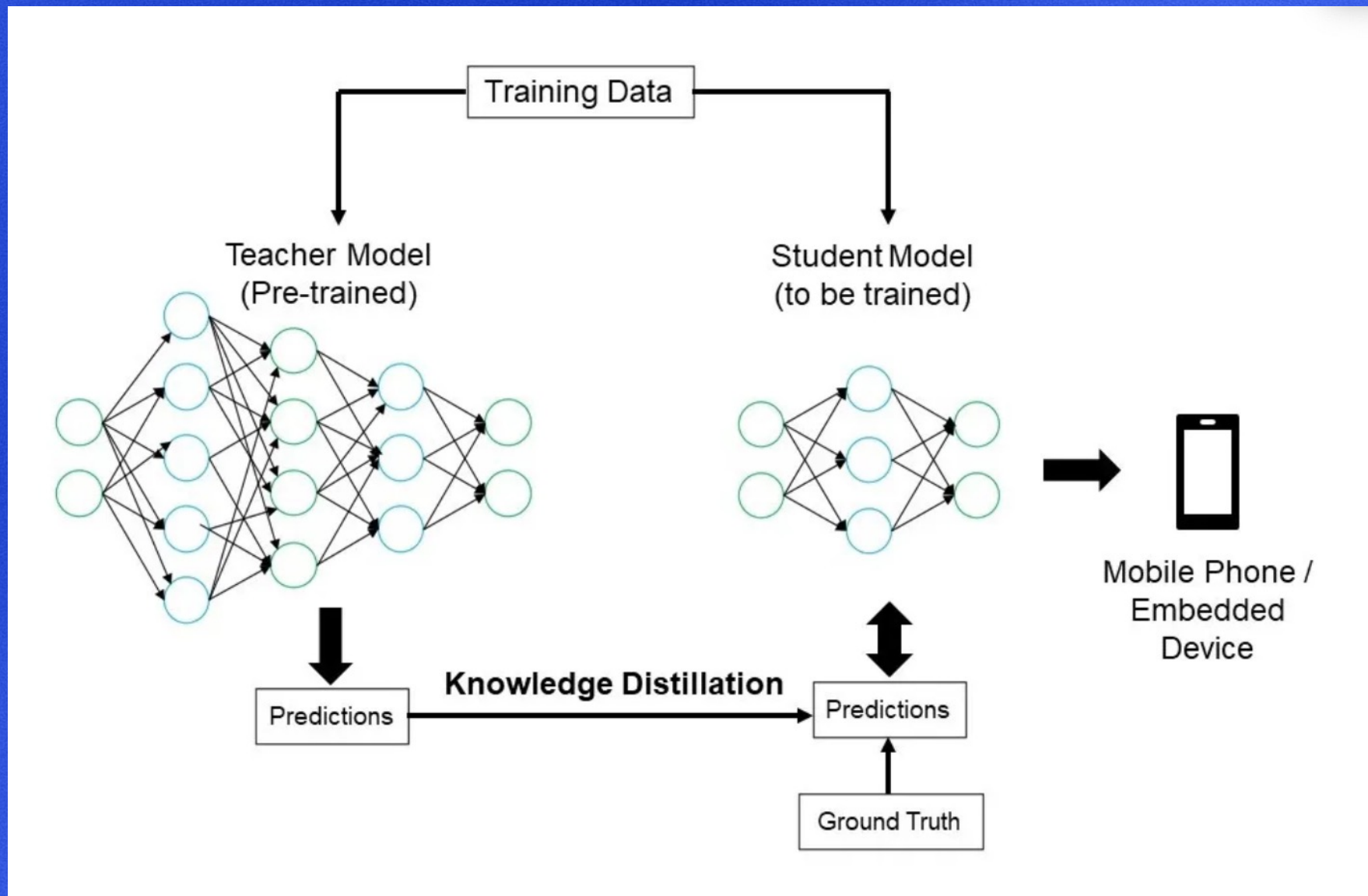
1. Multimodaliteettimallit yleistyvät



Zhang, Duzhen, et al. "Mm-llms: Recent advances in multimodal large language models." *arXiv preprint arXiv:2401.13601* (2024).

Keskeiset tekniset kehityssuunnat

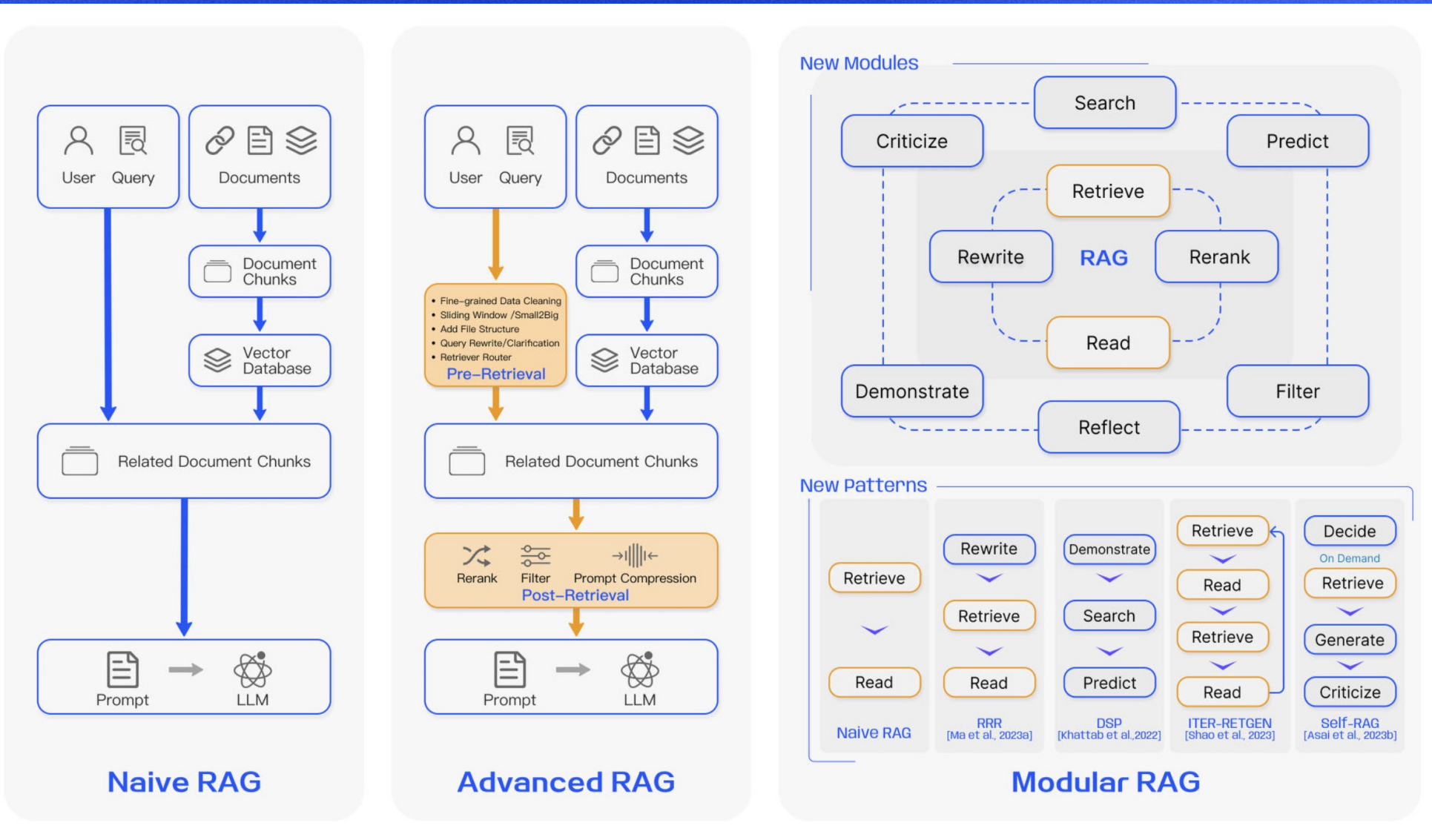
1. Multimodaleettimallit yleistyvät
2. Pienet kielimallit yleistyvät



Vijendra.1893. "Knowledge Distillation: Theory and End to End Case Study." Analytics Vidhya, 4 Jan. 2022, www.analyticsvidhya.com/blog/2022/01/knowledge-distillation-theory-and-end-to-end-case-study/.

Keskeiset tekniset kehityssuunnat

1. Multimodaleettimallit yleistyvät
2. Pienet kielimallit yleistyvät
3. RAG ja pidentyvät kontekstit tekevät kielimalleista hyödyllisempiä



Gao, Yunfan, et al. "Retrieval-augmented generation for large language models: A survey." *arXiv preprint arXiv:2312.10997* (2023).

Keskeiset tekniset kehityssuunnat

1. Multimodaleettimallit yleistyvät
2. Pienet kielimallit yleistyvät
3. RAG ja pidentyvät kontekstit tekevät kielimalleista hyödyllisempiä
4. Few shot learning ohittaa kielimallien hienosäätämisen

Few shot learning

Lonkalta

Zero shot learning

*Käännä englannista suomeksi:
cheese =>*

Kerrasta

One shot learning

*Käännä englannista suomeksi:
sea otter => merisaukko
cheese =>*

Kourallisesta

Few shot learning

*Käännä englannista suomeksi:
sea otter => merisaukko
peppermint => piparminttu
coffee => kahvi
juusto =>*

Keskeiset tekniset kehityssuunnat

1. Multimodaleettimallit yleistyvät
2. Pienet kielimallit yleistyvät
3. RAG ja pidentyvät kontekstit tekevät kielimalleista hyödyllisempiä
4. Few shot learning ohittaa kielimallien hienosäätämisen

KIELIMALLIEN TULEVAISUUS

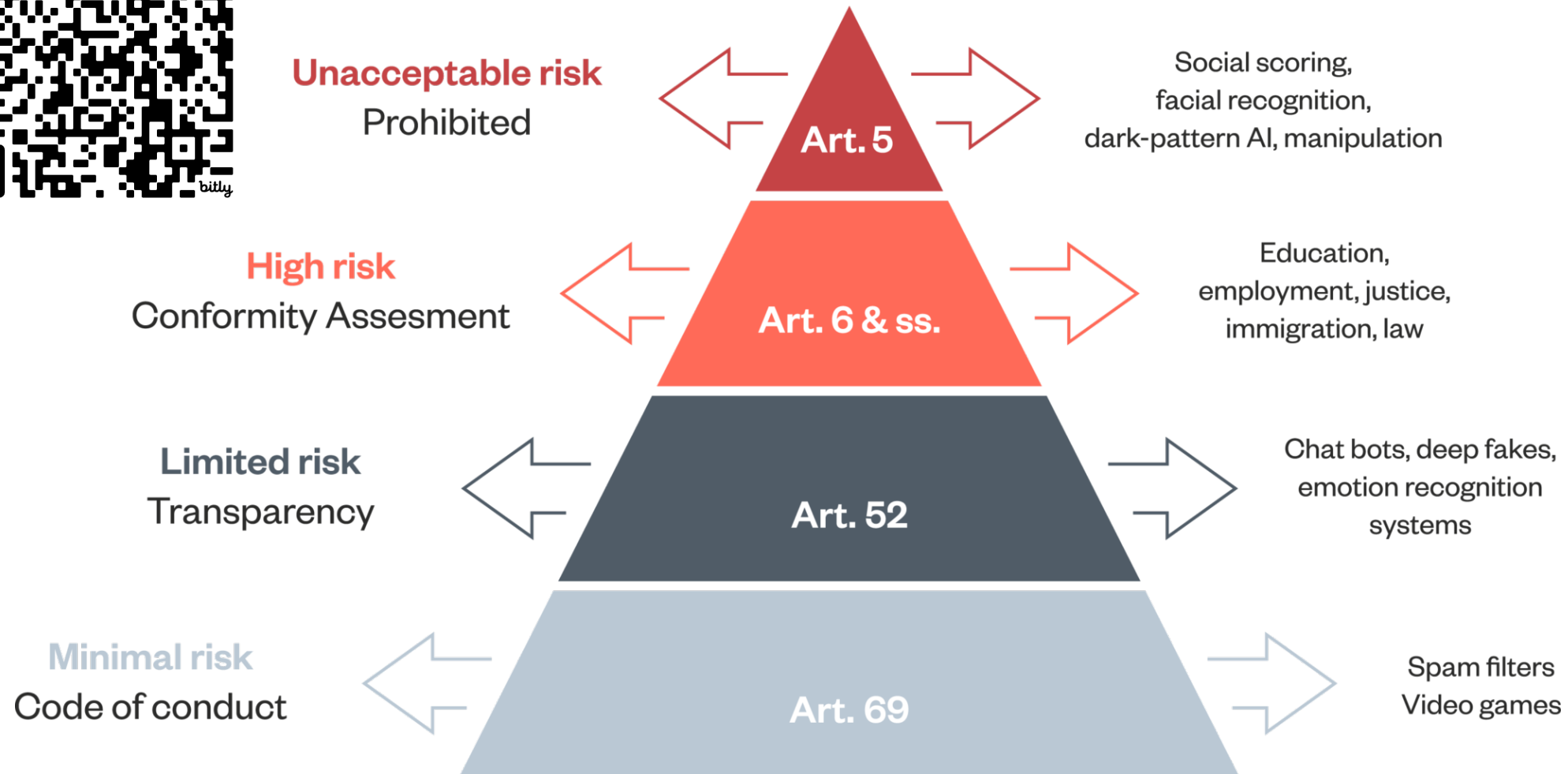
EU:n AI asetus

Kehittäjä (provider)

- Tekoälyjärjestelmän yleinen käyttötarkoitus
- Tekoälyjärjestelmän tunnetut rajoitukset
- Tunnetut, todennäköiset ja erityisen suuret riskit, joita saattaa esiintyä, ja toimenpiteet näiden riskien vähentämiseksi
- Yleiskatsaus tekoälyjärjestelmän kouluttamiseen käytetyistä tiedoista

Käyttäjä (deployer)

- Tarkoitus, johon käyttäjä aikoo tekoälyjärjestelmää käyttää
- Avoimuustoimenpiteet, mukaan lukien ilmoitukset tekoälyjärjestelmän vaikutuspiirissä oleville ihmisille
- Tunnetut, todennäköiset ja erityiset suuret riskit, joita saattaa esiintyä, ja toimenpiteet näiden riskien vähentämiseksi
- Käyttöönoton jälkeinen seuranta



"Expert Explainer: The EU AI Act Proposal." www.adalovelaceinstitute.org, www.adalovelaceinstitute.org/resource/eu-ai-act-explainer/.

MITEN ALKUUN KIELIMALLIEN HYÖDYNTÄMISESSÄ?

3 tärkeintä asiaa juuri nyt

1. Aina on hyvä hetki aloittaa kokeilut
2. Sääntelyn ja kustannusten huomiointi – viisas pääsee vähemmällä
3. Keskiössä on organisaation strategia kielimallien käytöstä

GenAI Center of Excellence – a Smartbi offering

With GenAI Expertise as a Service, you get our proven practices, templates and knowledge for your use through a monthly quota of hours – which you can focus in the most pressing focus area or even handle the entire solution delivery life-cycle:



Educate & Discover

Improve how business units understand the opportunities GenAI brings and actively support the identification of potential use cases.



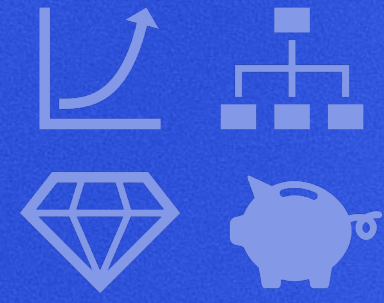
Define & Evaluate

Define and structurise the found use cases to better understand how GenAI should be applied and evaluate the business value and risk the potential solution has.



Plan & Pilot

Decide the most suitable technical approach for your tech stack preferences to plan and execute a proof-of-concept or pilot project of the solution



Implement & Scale-up

After reaching measurable business value in PoC or pilot project, increase the solution maturity to production-grade or scale the solution to wider audience

OTA YHTEYTTÄ



EMIL SIEVINEN

CHIEF DATA SCIENTIST

emil@smartbi.fi

+358 40 0781 213



EU AI ACT - podcast



smartbi