

Keeletehnoloogia Bürokratis

Merle Paula Videvik
Keeletehnoloogia tootejuht



Suur keelemudel (SKM)

- Plaanime kasutusele võtta kahes etapis:
 - Üldteadmuse jagamine – 2024 lõpuks
 - Asutuste avaliku info jagamine – 2025 I p.a.
- Eeldused:
 - Andmebaas automaatselt uuenevate andmetega
 - RAG (Retrieval Augmented Generation) süsteem
 - SKM, mis oskab eesti keelt, vastab kiiresti ja kvaliteetselt ning on mõistliku hinnaga

Testimine – www.buerokratt.ee

- Teemad: sotsiaalvaldkond ja eraisiku maksud
- Andmeallikad: eesti.ee artiklid ning seal viidatud veebilehed, Riigi Teataja
- Testperioodi piirangud:
 - Andmed on seisuga 30.05.2024 ja ei uuene
 - Ei saa esitada jätkuküsimusi
 - Ainult avalik info
 - Puuduvad viited algandmetele
- Ilmnenud probleemid:
 - Väike muutus küsimuse sõnastuses võib oluliselt muuta vastuse sisu
 - Veebilehtede sisu puhastamine on ajamahukas

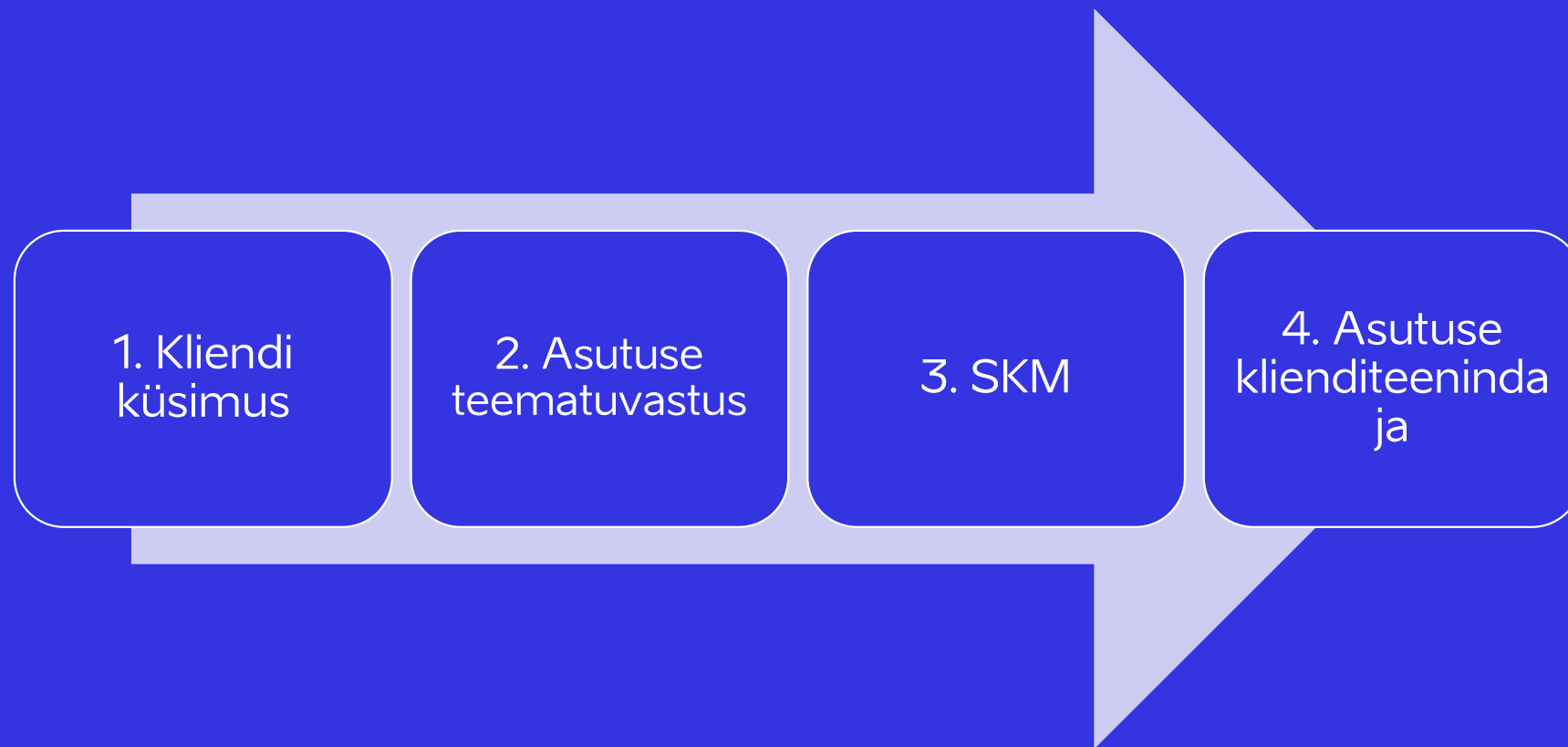
Üldteadmusbbaas

- Avaliku sektori valdkonnad — tervishoid, toetused, sotsiaalteenused, pensionid, transport, omand, töösuhted, keskkond, perekond, haridus jne
- Andmeallikad — eesti.ee artiklid ja seal viidatud veebilehed, Riigi Teataja
- Uuenevad andmed — perioodiline andmeallikate monitoorimine ning uuendamine
- Teabevärav — andmete hoiustamine

Tegevused keelemudeli kasutuselevõtuks

- Testimise kokkuvõte, vajalike muudatuste tegemine (august – september)
- Riskianalüüs (august – september)
- Bürokrati andmekaitse mõjuhinnaang (septembriks)
- Üldteadmusbaasi arendamine (oktoobriks)

Vestluste teekond SKM-iga



Klassifitseerija

1. Faas I — lokaalne e-kirjade klassifitseerija (arenduse lõpp august 2024)
2. Faas II — lokaalne vestlusandmete klassifitseerija (jätk — 2024 lõpp)
3. Faas III — globaalne asutustevaheline klassifitseerija (jätk - 2025)

Anonümiseerija

- Algne versioon olemas, sisend edasaiarendusteks olemas
- Vajalik muuhulgas klassifitseerija treeningandmete anonümiseerimiseks
- Kasutatav nii eraldiseisva tootena kui Bürokratis
- Edasiarendus (2024) ja Bürokrati integratsioon (2025)

Tänu!



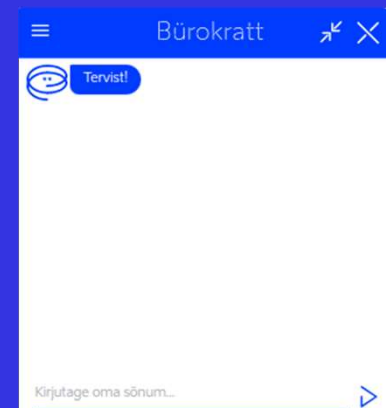
Bürokrati tulevik

Markko Liutkevičius
MõKt talituse juhataja

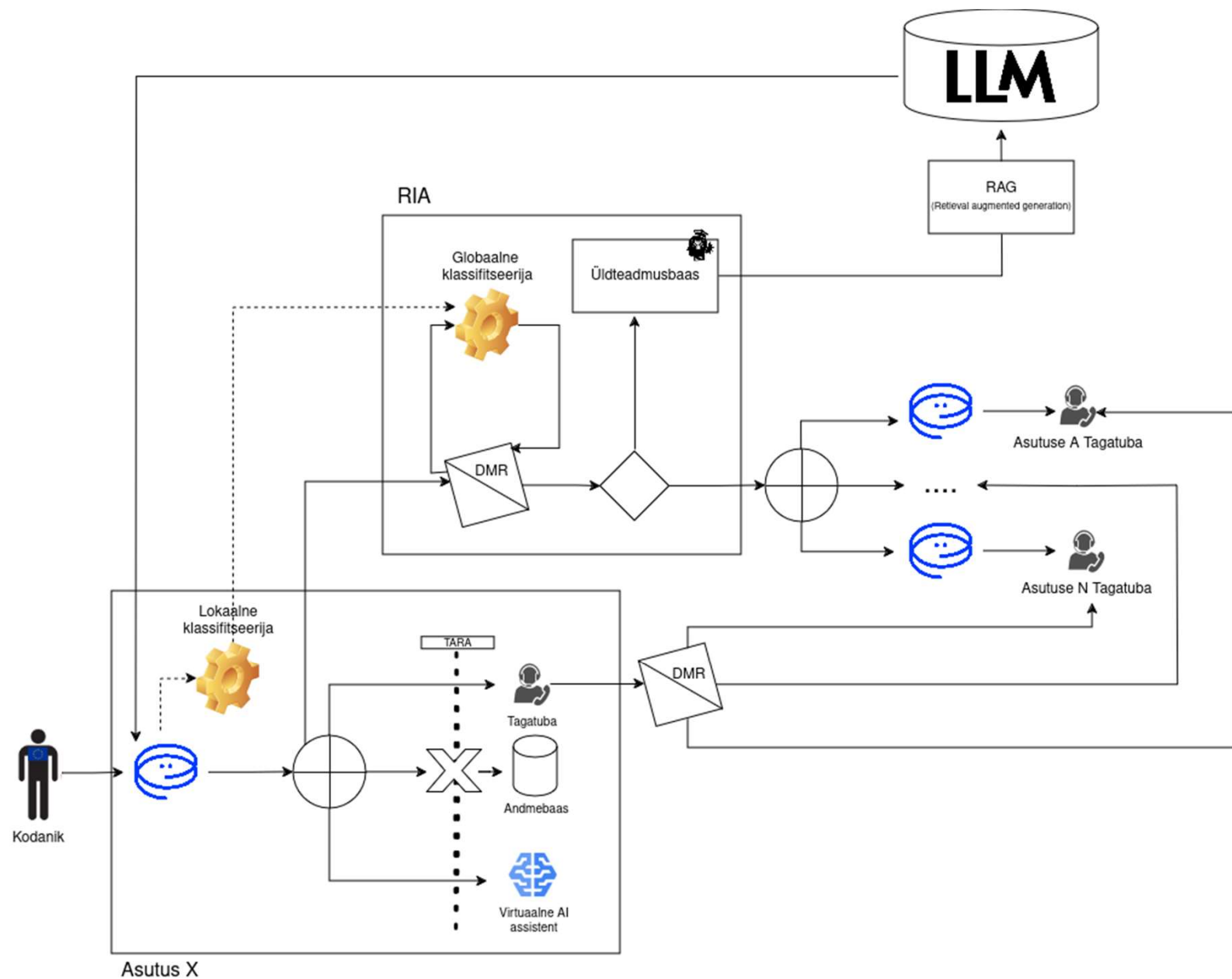


Visioon

„Võrgustik, mis kodaniku vaatepunktist toimib ühtse kanalina avalikele teenustele ja teabele ligipääsuks”



Bürokrati vundament



Bürokrati teekaart

| | VIII | IX | X | XI | XII | 2025 Q1 | 2025 Q2 | 2025 Q3 | 2025 Q4 |
|-------------------|--|---|--|--|---|--|--|--|---|
| Versioonid | 2.0 | 2.1 | 2.2 | 2.3 | 2.4 | 2.5 | | 3.0 | |
| Bürokratt | Andmete migreerimise lahendus. V2 esimese kliidini | Klientide üleviimine V2 peale, Multidomeeni lahendus | Isikustatud X-tee teenused, Jira liidestus | Teenuste mooduli tootestamine | Bürokrati keskne statistika avalikustamine | Vestluse suunamine teisele asutusele | Kavatsuse suunamine õige Bürokrati juurde. Riikliku postkasti teenusega liidestamine | Vestluse koopia saatmine e-mailile, Manuse saatmise võimalus | Esimene avaliku sektori soovitusüsteemi liidestamine Bürokratiga. |
| Keele-tehnoloogia | Riigisektori avalikku infot sisaldav suurkeelemudel avalikus testis (sotsiaalvaldkond) | Üldteadmusbaasi automatiseerimise eeltööd, hange ja arendused | Anonümiseerija edasiarendused | (Lokaalse) Klassifitseerija edasiarendused | Riigisektori avalikku infot sisaldav suurkeelemudel Bürokratiga liidestatud | Masintõlke lahendused | RAG tehnoloogia integratsioon Azure asemel lokaalseks. sh asutuste andmetega kasutus | Globaalne klassifitseerija, Speech-to text ja text-to-speech | |
| Vundament | Monitooring kliendi juures lahendus | XTR ehk X-tee päringute kiht | Turvis | Test driven development juurutamine | Keskne monitooring, DMR-i ehk BK võrgustikule aadressiraamatu valmimine | CentOps, GovSSO, End-to-end encryption | | | |

Kes maksab?

- Majutus
- SKM
- RAG

Tänu!

