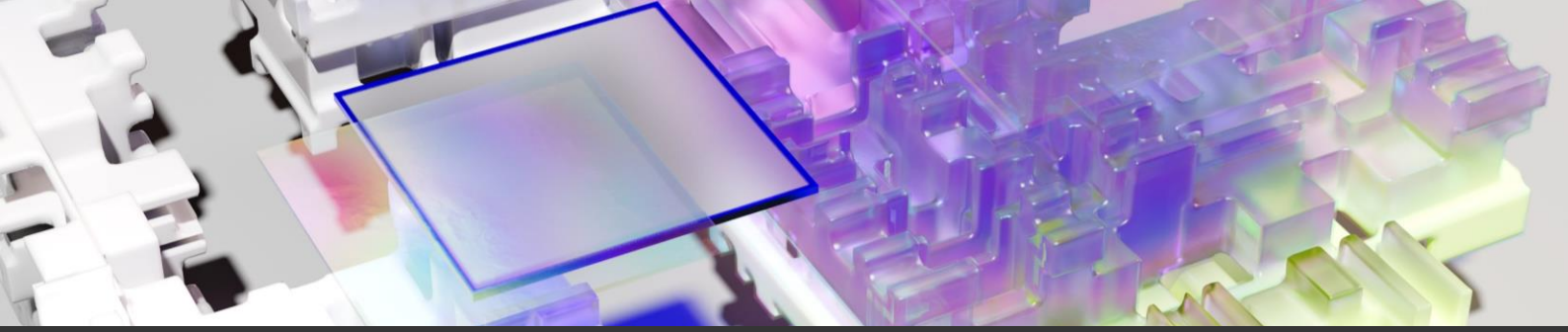


An AI-based implementation model for providing smart and user-centric services through Digital Gateway for Enterprises

Full official title: “Report on an AI-based implementation model for providing smart and user-centric public services and information through Estonia’s single digital contact point for enterprises”
within the project “Supporting productivity and competitiveness of Estonian SMEs through real-time economy and single contact point digital solutions”

Tallinn
2023

pwc

Republic of Estonia
Ministry of Economic Affairs and Communications

DISCLAIMER

This Report on "an AI-based implementation model for providing smart and user-centric public services and information through Estonia's single digital contact point for enterprises" was produced with the financial support of the European Union.

The opinions expressed in this document are the sole responsibility of the author and do not necessarily represent the official position of the European Union.

This project is funded by the European Union *via* the Technical Support Instrument, managed by the European Commission Directorate-General for Structural Reform Support.

Project reference: REFORM/SC2021/092 and Framework Contract SRSS/2018/01/FWC/002
Report provider: PwC
Financer: European Commission, DG REFORM
Beneficiary: Estonian Ministry of Economic Affairs and Communications

© European Commission 2023

Glossary

Term	Definition
AI	Artificial intelligence
AIOPs	Artificial intelligence operations
API services	Application interfaces services
ARIB (PRIA)	Agricultural Registers and Information Board (*Põllumajanduse Registrite ja Informatsiooni Amet*)
BERT	Bidirectional Encoder Representations from Transformers
BxT Agile framework	Business Customer Technology Agile framework
CFO	Chief Financial Officer
CMS	Content management system
CRIS (RIK)	Centre of Registers and Information Systems (*Registrite ja Infosüsteemide Keskus*)
DG REFORM	The European Commission's Directorate-General for Structural Reform Support
DGE	Digital Gateway for Enterprises
Digital ID	Digital identity
e-ETCB (e-MTA)	e-services environment for Tax and Customs Board (*Maksu- ja Tolliameti e-teenuste keskkond*)
EAS	Enterprise Estonia/ Estonian Business and Innovation Agency
EASME	Estonian Association of SMEs (*Eesti Väike- ja Keskmiste Ettevõtjate Assotsiatsioon*)
ECCI	Estonian Chamber of Commerce and Industry (*Eesti Kaubandus-Tööstuskoda*)
ECSA (EVS)	The Estonian Centre for Standardisation and Accreditation (*Eesti Standardimis- ja Akrediteerimiskeskus*)
EIC (KIK)	Environmental Investment Centre (*Keskkonnainvesteeringute Keskus*)
ELT/ETL	Extract load transform \ Extract transform load
EMCS	Excise Movement and Control System
EMTAK	The Estonian Classification of Economic Activities (*Eesti Majanduse Tegevusalade Klassifikaator*)
ETCB (MTA)	Tax and Customs Board (*Maksu- ja Tolliamet*)
EU	European Union
FAQ	Frequently asked questions
GDPR	General Data Protection Regulation
GPU	Graphics processing unit
GravCMS	Content management system
GridSearchCV	GridSearchCV is a technique for finding the optimal parameter values from a given set of parameters in a grid. It is essentially a cross-validation technique.
IP rights	Intellectual property rights
ISA (RIA)	Information System Authority (*Riigi Infosüsteemi Amet*)
JWT token	Json web token
K-means++	k-means++ is an algorithm for choosing the initial values (or "seeds") for the k-means clustering algorithm.
kNN	k-nearest neighbours (k-NN) is a non-parametric supervised learning method used in machine learning.
ML	Machine learning
MLOps	Machine learning operations

Term	Definition
MVP	Minimum viable product
PCA	Principle Component Analysis
PDPA	Personal Data Protection Act
PIA	Public Information Act
PMO	Project management office
PoC	Proof of concept
PostgreSQL	It is a relational database management system emphasising extensibility and SQL compliance.
PwC	PricewaterhouseCoopers
RACI	RACI is an acronym used in the responsibility matrix derived from the four key responsibilities: responsible, accountable, consulted, and informed.
REA	Register of Economic Activities (*Majandustegevuse register*)
RfS	Request for Services
RMSE	Root means square error
RSS	Really Simple Syndication
SAO	In natural language processing, the subject–action–object (SAO) structure is used to convert unstructured textual data into structured textual data.
SDN	Software-driven network
SHAP	SHAP (SHapley Additive exPlanations) is a game theoretic approach to explain the output of any machine learning model.
SMEs	Small and medium-sized enterprises
SSSC (RTK)	The State Shared Service Centre (*Riigi Tugiteenuste Keskus*)
SQL MS	It is a relational database management system developed by Microsoft.
TARA	Governmental authentication service (*Riigi autentimisteenus*)
TED	Tender electronic daily
UI	User interface
UX	User experience
VMs	Virtual machines
XML	Extensible markup language
X-Road	A platform that allows secure and standardised data exchange between the state and the private sector[1] (in Estonian *X-tee*)

[1] ISA, X-tee, What is X-tee? www.x-tee.ee

Executive summary

Scope of the Project

This report has been developed within the Project carried out by PricewaterhouseCoopers (hereafter – PwC) on behalf of the DG REFORM, according to the specific contract No REFORM/SC2021/092 signed on December 1, 2021. The report covers the items required in the Request for Services (RfS).

This report covers Outcome 3 of this project - Validated best practice models.

Purpose of the Report

The report on "an AI-based implementation model for providing smart and user-centric public services and information through Estonia's single digital contact point for enterprises" is a part of the project "Supporting productivity and competitiveness of Estonian SMEs through Real-Time Economy and single contact point digital solutions".

The report consists of six chapters. Chapter 1 describes the scope of work and project methodology whereas the other chapters present the results achieved for different milestones of the project. The first milestone was to gather and analyse information to evaluate and select the potential AI use cases for the Digital Gateway for Enterprises (DGE). The second milestone was to select the most beneficial AI use cases. The third milestone was to prepare an AI implementation model for DGE infrastructure, and the fourth milestone was to outline a detailed practical plan to implement AI use cases in the DGE environment.

Chapter 2 explains the results achieved during the first milestone. It describes the details about data sources which were assessed to evaluate the AI use cases. Also, it explains the information about the current state assessment completed with respect to data availability, data quality, technical architecture, and governance structure of the state portal.

Chapters 3 and 4 explain the results achieved during the second milestone. The two most beneficial use cases were selected using the selection criteria. The selection criteria contained two main dimensions: "impact" and "ease of implementation". These two dimensions further consist of sub-criteria which assessed other aspects such as cost, benefit, implementation complexity, and technical infrastructure of the use cases.

Moreover, a survey was conducted to better understand how entrepreneurs perceived the usefulness of each of the use cases. The results were validated with organisations that represented entrepreneurs. The final selection of the use cases was made with the beneficiary.

Chapters 5 and 6 describe the results achieved during the third and fourth milestones. An AI implementation model was developed for the DGE infrastructure. Moreover, these chapters explain how to begin an implementation and what activities should be completed to develop AI models into the DGE's infrastructure.

For example, what AI models should be developed and how the AI models can be integrated with can be target architecture. Moreover, the target governance structure was defined which described how to manage the implementation. PwC Analytics and AI Transformation Framework, which consists of the following six dimensions was used to consider different aspects related to the AI implementation model:

1. Technology and Infrastructure
2. Data and Information
3. Organisation, Culture and Talent
4. Process and Integration
5. User-centered design
6. Legal

Furthermore, implementation roadmaps were developed for selected use cases. These roadmaps provide the details about practical development plans, risks, roles and responsibilities, and cost estimations for the selected use cases.

Main takeaways

The main takeaways from the report were as follows:

- Currently, the Digital Gateway for Enterprise has low AI readiness because of the fragmented data sources and data quality.
- During the project, thirty-plus different data sources and public sector services were investigated, and only 20% of them offer interfaces which can be used to gather data. Moreover, the data quality was assessed, and it was identified that 25–30% of the data sources have medium-level data quality, and the rest of data sources contained low quality data.
- Initially, twenty-plus AI use cases were identified, and 60% of them, which were most aligned with the DGE context, were taken into consideration for further assessment. The use cases were derived from the user needs compiled in Deliverable 3.
- The two most beneficial AI use cases were selected after rigorous evaluation. These use cases are named 1) "Personalised AI-based recommendations to improve DGE services" and 2) “Smart suggestions to entrepreneurs related to deadlines and obligations via the DGE portal".
- The selected use cases address multiple user needs. For example, a) users sometimes miss deadlines and want to be aware of their obligations, b) users want to receive relevant information based on their business events or current situation c) users need access to all relevant public services from one centralised location in a personalised way, d) users need to have an overview of upcoming obligations and opportunities.
- A survey of the entrepreneurs was conducted to validate the internal use cases' selection done by PwC team and beneficiary. The survey was shared with approx. 2,000 entrepreneurs and only 31 responses were received. Approximately 95% of entrepreneurs out of 31 selected the same two use cases as the most beneficial ones.
- The interviews with the Estonian Chamber of Commerce and Industry and the Estonian Association of SMEs were conducted to validate the internal use cases' selection done by PwC team and beneficiary. The representatives of these organizations scored all the use cases and approximately 90% of the results were similar to the internal evaluation done by PwC team and beneficiary.
- The two different implementation plans with practical steps were developed for selected use cases. These plans described all the activities to develop AI models and minimum viable products (MVP). Moreover, these plans also consist of cost and effort estimations.

Lühikokkuvõte

Projekti ulatus

Käesolev aruanne on koostatud Euroopa Komisjoni struktuurireformide toe peadirektoraadi tellimusel PricewaterhouseCoopers (edaspidi PwC) poolt läbiviidud Projekti raames vastavalt 1. detsembril 2021. aastal allkirjastatud lepingule nr REFORM/SC2021/092. Aruande koostamisel on lähtutud Projekti lähteülesandes esitatud nõuetest.

Aruandes kajastatakse ainult Projekti kolmanda tulemiga piiritletud teemasid – valideeritud parima praktika mudelid.

Aruande eesmärk

Käesolev aruanne "tehisintellekti rakendamise mudelist nutikate ja kasutajakeksete teenuste osutamiseks ettevõtjatele mõeldud ühtse digivärava kaudu“ on projekti „Eesti väikese ja keskmise suurusega ettevõtete tootlikkuse ja konkurentsivõime toetamine reaalajamajanduse ja ühtse kontaktpunkti digilahenduste kaudu“ üks tulemitest.

Aruanne koosneb kuuest peatükist, millest esimeses kirjeldatakse töö ulatust ja metoodikat, ning teistes käsitletakse seatud eesmärkide täitmist ning tulemeid. Projekti esimene eesmärk oli koguda ja analüüsida teavet, et hinnata ning valida kõige suurema potentsiaaliga tehisintellekti kasutusjuhud ettevõtja digivärava keskkonnas rakendamiseks. Teiseks, valida ettevõtja digivärava jaoks kõige kasulikumad tehisintellekti kasutusjuhud. Kolmandaks, luua tehisintellekti rakendamise mudel, mis sobituks ettevõtja digivärava infrastruktuuriga. Neljandaks, töötada välja üksikasjalik praktiline plaan tehisintellekti kasutusjuhtude rakendamiseks ettevõtja digiväravas.

Aruande teine peatükk annab ülevaate tegevustest ja tulemustest, mis viidi ellu esimese seatud eesmärgi saavutamiseks. Peatükis antakse ülevaade tehisintellekti kasutusjuhtudega seotud andmeallikatest ning andmeallikate hindamise tulemustest käsitledes andmete kättesaadavust, kvaliteeti, tehnilist arhitektuuri ning ettevõtja digivärava juhtimisstruktuuri.

Aruande kolmas ja neljas peatükk annavad ülevaate teiseks seatud eesmärgi tegevustest ja tulemustest. Valiti välja kaks kõige kasulikumat tehisintellekti kasutusjuhtu valikukriteeriumite alusel. Valikukriteeriumid jagunevad kaheks peamiseks dimensiooniks – „mõju“ ning „rakendamise lihtsus“, mille abil on võimalik tuvastada kasutusjuhud, mis omavad kõige suuremat mõju ettevõtjatele ning mida on kõige lihtsam digiväravas rakendada. Nimetatud dimensioonid jagunevad veel omakorda alamkriteeriumiteks, mille alusel hinnati kasutusjuhtude kulu, kasulikkust, rakendamise keerukust ning tehnilist infrastruktuuri.

Lisaks viidi läbi küsitlus, et paremini mõista, kuidas ettevõtjad tajuvad kasutusjuhtude kasulikkust. Tulemusi valideeriti hiljem ettevõtteid esindavate katusorganisatsioonidega. Lõpliku valik kasutusjuhtude osas tehti koostöös Majandus- ja Kommunikatsiooniministeeriumiga.

Aruande viies ja kuues peatükk kirjeldavad kolmanda ja neljanda eesmärgi täitmist ja tulemusi. Välja töötati tehisintellekti rakendamise mudel ettevõtja digivärava jaoks, mis kirjeldab, kuidas integreerida tehisintellekti ettevõtja digiväravas võttes arvesse portaali infrastruktuuri ning teisi olulisi aspekte. Lisaks kirjeldatakse tehisintellekti lahendusi toetavaid toimimise tavasid ning tuleviku arhitektuuri ja haldamise põhimõtteid. PwC analüüsi ja tehisintellekti kujundamise raamistikku kasutati tehisintellekti rakendamise mudeli analüüsimiseks ja kirjeldamiseks, ning see koosneb kuuest kategooriast:

1. Tehnoloogia ja infrastruktuur
2. Andmed
3. Organisatsioon, kultuur ja inimesed
4. Protsessid ja integratsioon
5. Kasutajakesksus
6. Õiguslik taust

Kõige kasulikumate kasutusjuhtude kohta koostati teekaardid, mis annavad praktilised juhised lahenduste rakendamise planeerimiseks, sh riskide hindamiseks, rollide ning vastutuste määramiseks ning võimaliku maksumuse hindamiseks.

Peamised järeldused

Aruande peamised järeldused on järgmised:

- Ettevõtja digivärava valmisolek tehisintellekti rakendamiseks on madal, seda ennekõike killustunud andmeallikate ning andmekvaliteedi tõttu.
- Projekti käigus uuriti enam kui 30 erinevat andmeallikat ning avalikku teenust. Hinnanguliselt 20%-l uuritavatest allikatest on sobivad liidesed andmete pärimiseks, ning umbes 25–30% allika puhul on andmekvaliteet keskmisel tasemel. Ülejäänud allikate puhul hinnati andmekvaliteet madalaks.
- Projekti varasemas aruandes (D3, *deliverable 3*) kirjeldatud kasutajate vajaduste põhjal tuletati enam kui 20 potentsiaalset tehisintellekti kasutusjuhtu. Neist 60% sobitus ettevõtja digivärava kontseptsiooniga kõige enam ning neid käsitleti lähemalt töö järgnevates etappides.
- Pärast põhjalikku hindamist valiti välja kaks kõige kasulikumat tehisintellekti kasutusjuhtu: „Personaliseeritud soovitused vastavalt ettevõtja viimastele tegevustele ettevõtja digiväravas“ ning „Automaatsed teavitused ettevõttele avaliku sektori poolt seatud kohustuste ning tähtaegade kohta“.
- Valitud kasutusjuhud käsitlevad mitmeid kasutajate vajadusi, näiteks a) ettevõtjad tahtmatult mõnikord unustavad avaliku sektori poolt seatud kohustuste tähtaegasid ning tahaksid seetõttu olla paremini informeeritud neile seatud kohustustest; b) ettevõtjad soovivad saada asjakohaseid soovitusi nendega seotud ärisündmuste ning -tegevuste kohta; c) ettevõtjad soovivad ligipääsu kõikidele asjakohastele avalikele teenustele ning teabele ühest kanalist, eelduslikult personaliseeritud lähenemisega; d) ettevõtjad soovivad saada ülevaadet nii kohustustest kui ka võimalustest.
- Pärast esmast töögrupisisest kasutusjuhtude hindamist saadeti välja küsimustik ligikaudu 2 000 ettevõttele, et saada nende tagasisidet valitud kasutusjuhtudele. Küsimustikule vastas 31 ettevõtet ning tulemused ühtisid 95% ulatuses esialgsete hinnangutega.
- Ettevõtjatelt saadud tagasiside valideeriti Eesti Kaubandus-Tööstuskoja ning Eesti Väike- ja Keskmiste Ettevõtjate Assotsiatsiooni esindajatega. Intervjuudelt saadud hinnangud kattusid 90% ulatuses esialgsete hinnangutega.
- Kõige kasulikumate kasutusjuhtude jaoks töötati välja implementeerimise mudelid koos praktiliste teekaartidega, mis annavad suunised vajalikeks tegevusteks tehisintellekti mudelite ning minimaalsete töötavate lahenduste (*minimum viable product*) loomiseks. Lisaks käsitlevad need hinnanguid töömaksumuse osas.

Table of Contents

1. SCOPE OF WORK, PROJECT METHODOLOGY 10

1.1 PROJECT TIMELINE FROM DIGITAL GATEWAY FOR ENTERPRISES STREAM PERSPECTIVE 10

2. INFORMATION GATHERING AND DOCUMENTS ANALYSIS 12

2.1 METHODOLOGY FOR INFORMATION GATHERING AND ANALYSIS 12
2.2 KEY FINDINGS ON INFORMATION AND DOCUMENTS ANALYSIS 13
2.3 CURRENT STATE ASSESSMENT 14

3. SELECTION CRITERIA FOR DGE AI USE CASES 21

3.1 METHODOLOGY OF SELECTION CRITERIA MEASURES 21
3.1.1 MEASURE ESTIMATED IMPACT 21
3.1.2 MEASURE ESTIMATED EASE OF IMPLEMENTATION 24
3.2 METHODOLOGY OF SELECTION CRITERIA SCALES 25
3.2.1 METHOD TO SCORE AI USE CASES 25

4. PREPARATION AND EVALUATION OF THE DGE AI USE CASES 27

4.1 PREPARE A LIST OF AI USE CASES 27
4.2 PREPARE DATA SOURCES LIST FOR AI USE CASES 29
4.2.1 LIST OF POTENTIAL DATA REQUIRED 29
4.3 EVALUATION OF DGE AI USE CASES 30
4.3.1 SCORES OF DGE AI USE CASES 30
4.4 VALIDATION OF THE EVALUATION RESULTS 31
4.5 SELECTED DGE AI USE CASES 34

5. IMPLEMENTATION OF AI IN THE DIGITAL GATEWAY FOR ENTERPRISES 36

5.1 METHODOLOGY FRAMEWORK: PWC ANALYTICS & AI TRANSFORMATION FRAMEWORK 36
5.1.1 TECHNOLOGY AND INFRASTRUCTURE 36
5.1.2 DATA AND INFORMATION 37
5.1.3 ORGANISATION, CULTURE & TALENT 38
5.1.4 PROCESS AND INTEGRATION 39
5.1.5 USER-CENTRIC DESIGN 40
5.1.6 LEGAL 40

6. IMPLEMENTATION ROADMAP FOR SELECTED AI USE CASES 42

6.1 IMPLEMENTATION ROADMAPS FOR DGE SELECTED USE CASES 42
6.1.1 USE CASE "PERSONALISED RECOMMENDATIONS" 43
6.1.2 USE CASE "SMART SUGGESTIONS FOR DEADLINES AND OBLIGATIONS" 48
6.2 ANNEX 1. DETAIL EVALUATION OF THE USE CASES 54
6.3 ANNEX 2. SURVEY FOR THE ENTREPRENEURS 58

1. Scope of work, project methodology

The current chapter introduces the project and its timeline, as well as the methodology of the activities carried out in the report.

1.1 Project timeline from Digital Gateway for Enterprises stream perspective

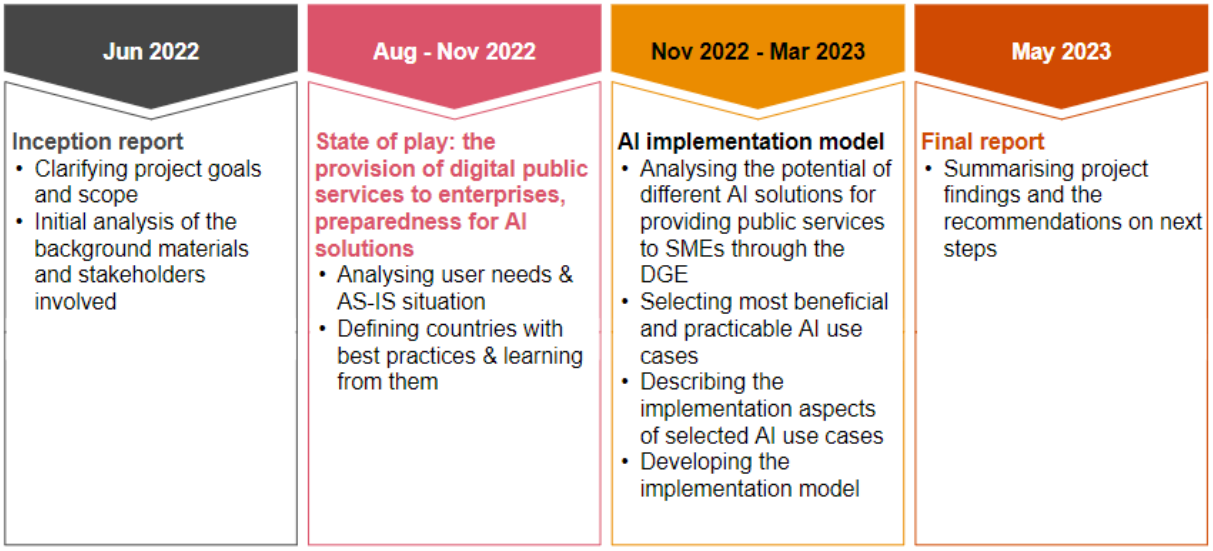
The main objective of the Digital Gateway for Enterprises is to improve the efficiency and the quality of public service provision and reduce the administrative burden on entrepreneurs. The project is implemented by the Estonian Ministry of Economic Affairs and Communications.

The purpose of the deliverable is to provide a report on the analysis and selection of innovative AI solutions for Estonia's Digital Gateway for Enterprises. Moreover, it describes the implementation model for selected AI use cases which covers practical aspects such as roadmap, resources and effort required, cost estimation, and risks. The outcome of the report is an implementation model which will be used to proceed further with the development of AI solutions in the Estonian Digital Gateway for Enterprises.

The report provides the list of AI use cases derived based on the user needs which were identified in Deliverable 3. The report describes the information analysis and data assessment carried out and explains the details of the evaluation and selection of the most beneficial AI use cases. Furthermore, an AI implementation model with roadmaps is described for the selected AI use cases, which outlines the detailed plan to begin the development of selected AI solutions in Estonia's single digital contact point for enterprises.

The project's expected duration is from June 2022 to May 2023. Please refer to the project plan in Figure 1 for specific DGE deliverables. Project milestones, aligned with the beneficiary, consist of the state of play analysis together with best practice analysis and creating an AI implementation model for the most beneficial use cases.

Figure 1. Project outline/components on DGE



The project methodology includes steps of gathering input, stakeholder engagement and expanding their contributions, preparing and conducting workshops, assessing the current state, defining the target state with selected AI solutions, and presenting deliverables.

Please refer to the methodology outline in Table 1.

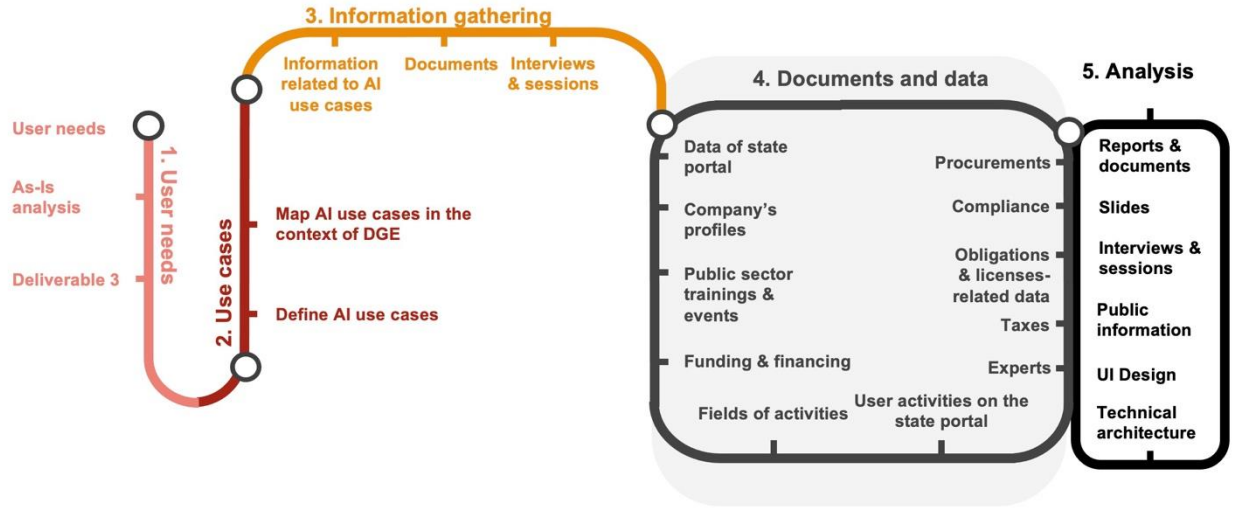
Table 1. Methodology

Deliverables	Methodology elements
• Deliverable 3: Report on the current situation in Estonia regarding the provision of digital public services to enterprises, preparedness for the use of AI solutions in public service provision, and the best practice study.	
State of play: the provision of digital public services to enterprises, preparedness for AI solutions	• Gather relevant DGE documents and confirm selections. • Perform document analysis. • Gather information and insights through workshops with stakeholders. • Choose selection criteria and confirm final country selection results. • Gather relevant documents for best practice analysis. • Perform best practice analysis. • Present deliverable to DG REFORM and beneficiary.
• Deliverable 7: Report on an AI-based implementation model for providing smart and user-centered public services and information through Estonian Single Digital Contact Point for Enterprises.	
AI-based implementation model	• Gather and analyze documents, data, and best practices. • Prepare, present, and confirm intermediate results with the beneficiary. • Select the most appropriate (beneficial and feasible) AI use cases with stakeholders. • Develop an AI implementation model.

2. Information gathering and documents analysis

This chapter explains the analysis of the information and documents gathered during deliverable D7. To complete the analysis, the PwC team followed an agile and iterative way to assess the information until the end. It helped the overall process to become flexible and scalable. The key elements of the process are illustrated in Figure 2 below.

Figure 2. Overview of the process to gather and analyse information



2.1 Methodology for information gathering and analysis

User needs and AS-IS analysis

User needs were gathered and identified during the preparation of the D3 AS-IS analysis report[2], which were used as a baseline for the information search. The analysis and the identified user needs provided a high-level overview of the entrepreneurs' requirements and analysis of various aspects of the DGE portal. The AS-IS analysis assessed high-level existing data and infrastructure elements.

The user needs explained in the deliverable D3 report mainly related to providing recommendations and suggestions for different business events, compliances and obligations. Some user needs were case-specific such as procurement-related suggestions and providing language support as not all entrepreneurs in Estonia speak Estonian.

The AS-IS analysis report also documented several aspects related to the DGE environment. For example, aspects related to organisation, process and integration, data, technology and infrastructure were covered. However, there were some elements linked to DGE AI use cases and implementation models which were further analysed during the D7, e.g., current technical architecture and data sources with underlying data systems.

[2] DGE D3 State of play report

Use Cases

The AI use cases were identified with respect to user needs and DGE context. The process was iterative, and multiple discussions with the beneficiary were held. Initially, the list of use cases was prepared which enlisted broader ideas because most of the identified user needs were captured as broader topics. During the process, the use cases were defined with the DGE context in mind and the ideas of use cases were narrowed down to gather substantial information.

Moreover, the topics related to the evaluation of the use cases were also discussed in parallel. The aim was to reduce the risk of producing invalid use cases that were difficult to evaluate. The factors and criteria to evaluate the use cases also contributed positively to select the most beneficial use cases.

For example, the use case which was addressing a user need “user needs to understand the language in which the information is provided to them (i.e. simple language)” was marked as less relevant. The reasons were the following: a) the difficulty in verifying the text (simple language) generated by AI especially text related to legislation, b) the risk that misinformation could be passed down to the entrepreneurs through the DGE portal.

Information gathering

The information and documents were gathered during the interviews conducted with data analyst and data scientist working on the state portal system. The main objectives of these interviews were to gather an up-to-date information about the current system. Moreover, several sessions were done with the beneficiary to analyse and evaluate the use cases. Different public services profiles e.g., tax and custom board and their data availability models were also assessed. Furthermore, the documents related to information architecture, business events and UI design of state portal were also analysed to draw out an understanding of existing architecture.

During the initial information assessments, the list of data categories was defined as shown in step 4 of Figure 2. The documents and data related to each category were assessed and mapped out into data availability, data quality and data accessibility. It was identified that these data categories were mainly associated with use case management, especially when evaluating the use cases with respect to data availability and readiness.

Documents and data

The documents and data specific to each category were assessed and key findings were mapped out. First, the underlying data source systems were identified to assess how the data can be extracted from the data sources. For example, some of the data sources supported API services and interfaces that could be used to gather data automatically, e.g., open data portal[3] provides the API services.

However, most of the data sources provided links to download data in pdf format. Different public services profiles and public registers such as e-Business register, registers related to of economic activities, and public procurement registers were also assessed from the data availability and quality perspective.

Analysis

As a result of analysis, all the information and documents gathered from the identified data sources are shown in Figure 4. Different reports related to DGE's current state and UI designs were also analysed to align the understanding of the current state and define the target state with respect to AI use cases.

2.2 Key findings on information and documents analysis

Figure 3 presents an overview of key findings related to the main data categories and topics that were identified from user needs and use cases. Each data category is also described with underlying main data sources and services.

[3] https://avaandmed.ariregister.rik.ee/en

Figure 3. Overview of the key findings related to different public sector data sources

State's portal data	Company's profile data	Taxes	Compliance	Fields of activities
State's portal (eesti.ee) contains different types of data sources, for example, related to user activities and sessions on the portal stored in **PostgreSQL**. Though it is on a high level and not qualitative and granular. Another data source stored in the state's portal system is related to the content, which is stored in **GravCMS**.	Company's profile data is maintained and stored mainly in the **e-Business Register**, for example, company-related public data can be viewed as a single query. The **open data portal** is also a source to retrieve company's data through e-Business Register, e.g. detailed company data query. Also, **Statistics Estonia's** database contains stats about companies.	Taxation-related content is stored in different systems, for example, the **articles-related to taxes** which can be accessed through **x-Road service**. Moreover, the **Tax and Customs Board (ETCB)** offers public data services which can provide tax-related information through **public data queries**, for example, payment on taxes or payments on tax liabilities.	Compliance-related data is spread across different levels for each field of activities. It is mostly available in the form of unstructured text in different official principles, journals, notes, and documents. These documents are mostly stored and available in **GravCMS**, **e-ETCB**, **EVS**, **Draft Act Information System**, and **State Legislation Documents** systems.	Data related to different fields of activities is available in different systems and mostly available in unstructured and textual formats. The state's portal (eesti.ee) contains information about field of activities which can be the acquired through **x-Road service** or other systems, e.g. from **Centre of Registers and Information Systems' (CRIS)** systems.
Licenses & obligations	**Fundings & financing**	**Procurements**	**Trainings & events**	**Public sector experts**
Licenses and notices data is present mostly in the form of clauses and text. The main systems to gather data for different licenses and notices related to economic activities are **EMTAK search (CRIS), REA, State Legislation Documents,** and **GravCMS** (which actually contains curated information about licenses & notices for activities)	The data sources related to financing opportunities for enterprises offered by the public sector are available in multiple systems and mostly available in the form of text. The main data sources are different CMS systems, e.g. **RTK.ee**, **PRIA.ee**, **KIK.ee, Draft Act Information System**. Also, the state's portal (eesti.ee) data system (**GravCMS**) contains content information.	The public procurements related data are mostly available in different registers, mainly in **Estonian Public Procurement Register**. Some of the data is also available and fetched via an **open data portal**. There are also other data sources which can be utilised to gather tenders information for EU-level, e.g. tender electronic daily (**TED**)	Public sector offers a wide range of training and available data sources are distributed among different institutions. The important data related to training and events are mostly in an unstructured format. Some of the potential data sources can be **e-ETCB**, **visiidid.ee**, **EAS**, **arenduskeskused.ee**, and **tootukassa.ee**.	Information related to public sector experts is **partially** available from **Riigitöötaja iseteenindusportaal**. The information is mostly available in an unstructured format and utilised by several public institutions. The other data source can be the X-Road **service** to gather information related to jobs.

2.3 Current state assessment

Data assessment

During the several workshop sessions and interviews, the following data systems were identified as potential sources which can be used in AI modelling and solution development. The assessment contained findings related to information availability and quality. Figure 4 presents an overview of the data assessment.

The PwC team conducted the interviews and workshops to complete the data assessment, especially the interviews with data analyst and data scientist were conducted to gather up-to-date information-related to the databases used in the state portal system. Furthermore, the slides[4] and reports[5] related to the information architecture and data architectures were thoroughly reviewed. These reports outlined some of the data sources and detailed architecture. However, the information related to data availability was not directly addressed especially with respect to AI solutions.

The first step to developing AI solutions was to explore the datasets regardless of the data formats. AI algorithms were mostly trained on datasets which were fragmented and stored in the form of unstructured text. During the data assessment, the PwC team assessed the potential data sources that could provide some sort of documents or raw information in the form of text or tables.

According to the data categories mentioned above in Figure 3, information related to various data sources and other public services profiles was gathered and discussed with the beneficiary. All information gathered on potential data sources and other public service profiles was assessed particularly in terms of data availability and data quality. These two aspects were the most critical ones with respect to developing AI solutions. Data quality is assessed using information gathered during interviews and utilising the data quality dimensions for machine learning models. The focus was to consider the data quality aspects for machine learning models[6], which are listed below:

- data completeness: It assessed if the required data is present.
- data uniqueness: It assessed if the data features are unique. It also relates to data preparedness for machine learning models development perspective.

[4] Riigiportaali analüütika tutvustus_07.09.2022.pptx

[5] Riigiportaali eesti.ee kasutatavuse analüütika, Nortal, 2021

[6] Data quality considerations for machine learning models

- data validity: It assessed if the data is valid for machine learning development purpose.

Moreover, the scales of high (H), medium (M), and low (L) were used to complete the data assessments as shown in Figure 4. The following points describe the explanation behind these scales:

- High (H): If the data was available through API interfaces and its quality was good, experiments could begin.
- Medium (M): If the data was available through manual work, e.g., downloads, and its quality was medium.
- Low (L): If the data could not be gathered even manually and its quality was low or unknown.

Below are the lists of potential data sources. However, the lists are not exhaustive.

Figure 4. Overview of the data assessment of different systems and services

Data	Data Sources	Description	Data Availability	Data Quality
State's portal data	GravCMS	It is an article repository which contains all the content that is available on the state's portal.	H	M
	PostgreSQL	It contains the user sessions and activities-related data of the state's portal.	H	L
	x-Road service	It's a data exchange platform which supports secure data transition through REST APIs.	H	M
Company's profile data	Open data portal	It provides business public data in a readable format through REST APIs.	H	M
	e-Business Register	The e-Business Register is the official portal of the Estonian state that includes the data of all legal entities.	M	M
Taxes	e-ETCB	e-ETCB supports inquiries, where public data from the register of taxable persons can be gathered.	H	L
	Open data portal	Data from open data portals related to the management of state revenues, the implementation of national tax and customs policy, and the protection of society and legal economic activity.	H	L
	x-Road service	x-Road service to fetch data related to taxes via REST APIs.	H	L
Compliance	EVS	It's a non-profit organisation supported by state institutes to provide different information and collection services mainly into standardizations.	L	L
	State Legislation Documents	It is a platform which provides consolidated Estonian texts on Estonian legislation and its English translations.	L	M
	e-ETCB	Data from Tax and Customs Board related to legal compliances.	H	L
	Draft Act Information System	It is an information system where draft acts are coordinated between institutions.	M	L
Fields of activities	Centre of Registers and Information Systems	It is an information service centre for various registers where different fields of activities-related requirements can be searched and gathered.	M	L
	x-Road service	Data related to fields of activities from x-Road service.	H	L
License & obligations	Centre of Registers and Information Systems	It is an information service centre for various registers where different fields of activities-related requirements can be searched and gathered.	M	L
	MTR	It is a portal of economic activities register which provide functionalities to search data related to licenses and notices. Also, there are about ten other similar registers in Estonia, which can be further analysed to gather data related to licenses.	L	M
	State Legislation Documents	It is a platform which provides consolidated Estonian texts on Estonian legislation and its English translations.	M	M
Fundings & financing	RTK.ee	It's a national support services centre that contains data related to funding and programs	L	M
	PRIA.ee	It provides services to support rural development and information on related funds	M	L
	Draft Act Information System	It's an information system where draft acts are coordinated between institutions.	M	L
	KIK.ee	It is an environmental investment centre which provides information related to grants and funds	M	L
Procurements	Estonian Public Procurement Register	It is an public procurement service where information related to published procurements, notices, contract information can be found and might be retrieved via x-Tee service.	L	M
	Open data portal	Data related to procurement public data in a readable format through REST APIs.	H	M
	TED Europe	It's an EU-level tenders information service to provide details about public procurements across EU member states.	L	M
Trainings & events	visiidid.ee	It is service that provides an information and data related to foreign visits.	L	L
	EAS	It provides training and services on developing businesses and getting financing.	L	L
	e-ETCB	e-ETCB provides information on taxes training and programs.	H	L
	arenduskeskused.ee	It supports the development of businesses with consultations and training.	L	L
Public Sector experts	Riigitöötaja iseteenindusportaal	It's a public employee self service portal that provides the information flow from institutions to the state's financial, personnel and payroll system SAP.	L	L
	x-Road service	Data related to jobs and public sector experts via x-Road service and REST APIs. Moreover, information-related public sector experts can also be found on public sector web pages.	H	L

GravCMS: It is an article repository which contains all the content that is currently available on the state portal. It was identified during the interview with the data analyst that retrieving the data roughly takes two

weeks. However, there was a challenge in automatically updating the data.[7] This repository contained data from different sources, including those related to tax and financing, but if there were any changes in the original sources, it was not automatically reflected in the state portal CMS.[8] The content is already stored in a presentable way and the quality is identified as medium. Moreover, the new CMS system, ARVA is being planned to replace GravCMS in the future.[9]

PostgreSQL: It is a database used to store different types of information related to state portal activities. For example, the information related to users' sessions, website logs and information related to the mailbox. During the interviews, it was identified that the data quality is low with respect to user activities-related data. However, the data availability was high and extracting data from the database would roughly take one-two weeks.[10]

X-Road service: It is a data exchange platform that allows secure and standardised data exchange between the state and the private sector. There was a variety of systems and services available on the X-Road platform, and to access the platform, an organisation had to create a service and enable other organisations to use it. The X-Road service is already being used in state portal architecture to gather information about different organisations in a secure and encrypted way. Since the service was enabled within the current architecture, the data availability status is high. Moreover, during the interviews and discussions, it was identified that the data quality was medium because most organisations publish data in a usable format.[11]

e-Business Register: It is the official portal of the Estonian state that includes the data of all legal entities registered in Estonia in a single environment. Public information related to legal entities could be requested from the portal and there were various services offered by the e-Business Register. The portal allowed legal entities to perform different administrative tasks via the portal such as making changes to the data submitting petitions, documents and annual reports. During the assessment, it was identified that the data availability was medium because the e-Business Register provided information through an open data portal and EMTAK search. Moreover, the data quality also was of the medium level, as the open data portal contained data in a machine-readable and usable format.[12]

Open data portal: It is a platform which provides open data to be used by all organisations. The portal provided an interface to access public information related to legal entities registered in Estonia. There were various methods to gather the data, e.g., the data could be either downloaded or it could be retrieved using API services. The API services provided several interfaces to gather information. During the assessment, it was identified that interfaces related to detailed company data queries, queries for a list of company-related documents and company annual report queries were potentially quite useful to extract the company's profile data.

Moreover, the portal supported a single query interface that allowed cross-use and reuse of data. It is important to point out that API services (XML) and downloadable datasets contained only public data from the e-Business Register, which could be viewed before, but for a service fee. During the assessment, the data availability through the open data portal was identified as high due to a variety of features and the data quality was assessed as medium because the portal provided the data in a machine-readable format.[13]

Statistics Estonia's database: It is a national statistical portal that provides a wide range of information related to different sectors. The portal had features for requesting statistics and viewing data tables through databases. There were different channels to access the data: either directly using Statistics Estonia API interfaces or via X-road service. There were guidelines developed to use statistical database API or integrate reporting software such as Power BI. The interfaces were created to decrease the data

[7] Ettevõtte finantseerimise ärisündmuse analüüsi lõpparuanne

[8] Riigiportaali eesti.ee kasutatavuse analüütika, Nortal, 2021

[9] Ettevotjate-uhtse-kontaktpunkti-infotehnoloogilise-arhitektuuri-analuus-23.07.2021

[10] Riigiportaali analüütika tutvustus_07.09.2022.pptx

[11] https://www.emta.ee/en/business-client/e-services-training-courses/how-use-e-services/x-tee-services

[12] https://ariregister.rik.ee/eng

[13] https://avaandmed.ariregister.rik.ee/en

submission burden of enterprises and administrative burden by making it more convenient to submit data required by the state. During the assessment, the data availability was identified as medium and data quality was also discovered as medium due to the availability of different channels that supported data extraction in a useable format.[14]

e-ETCB: Estonian Tax and Customs Board (*Maksu- ja Tolliamet*) also provides services to gather public data related to people and entities. The information could be gathered through either public query features offered by the Tax and Customs Board[15] or through an open data portal where different datasets related to tax and customs were available[16]. During the data assessment, it was investigated that the available data from the open data portal might be useful but requires further analysis. The datasets from the open data portal provided information on various topics related to enterprise taxation. For example, these datasets provided information related to taxes paid by enterprises, turnover and number of employees, aggregate data for VAT returns, and local business statistics by municipalities.

EVS: It is a centre for standardisation and accreditation (*Standardimis- ja Akrediteerimiskeskus*) that provides different services to streamline information, monitoring and collection services online. There were guidelines provided by the service to ensure that the technical standards related to the product, process or service were fit for its purpose. The general principles involved aspects related to compatibility, interchangeability, variety control, safety, protection of the environment and product protection. The service offered to request and gather information and required submitting a form to access the information through the portal. During the assessment, it was identified that the current service did not offer any API service to gather data automatically and the quality of the information was also unknown and required further analysis.[17]

Draft Act Information System (Eelnõude infosüsteem): It is a platform where draft acts are coordinated between institutions, submitted to the government and the parliament, and public consultation takes place. There was a variety of documents available on the system such as national documents, draft European Union legislation and other documents related to the decision-making process of the European Union. The system provided interfaces where anyone could monitor the processing of drafts, find documents in the information system using a search, participate in public consultations and submit comments on the document under approval. During the assessment, it was discovered that the current portal did not offer any API service to gather data automatically. However, the service provided an interface to search for different legal drafts and documents that could be downloaded in pdf format, e.g. if someone was interested in finding a legal framework for accelerating the deployment of renewable energy for public consultation, the related pdf documents could be downloaded.[18]

State Legislation Documents (Riigi Teataja): It is a platform which provides consolidated Estonian texts on Estonian legislation and its English translations. The service offered an electronic interface for gathering information related to legal consolidated texts translated into English. However, it was explicitly mentioned that the translations were unofficial texts, they had no legal force, and no one was to rely on them in judicial or any other official proceedings. The purpose of translations was to support non-native entrepreneurs in finding legal acts in English. The services offered searches for all types of acts, such as the law of obligations. Documents relating to different acts could be downloaded in pdf or text format which can be utilised for analysis.[19]

Estonian Classification of Economic Activities (EMTAK): The Centre of Registers and Information Systems (CRIS) provides information and documents on different fields of activities via the EMTAK search portal. The portal offered information that included finding laws, and associated registers where enterprises can apply for activity licenses or submit notices of economic activity. The information related to notices and licenses associated with the field of activities could be downloaded in pdf format. However, the information format and structure were not suitable for an applied analysis. During the assessment, it

[14] https://www.stat.ee/en/statistikaamet/developers

[15] https://www.emta.ee/en/business-client/registration-business/inquiries-and-certificates/public-data-inquiries

[16] https://avaandmed.eesti.ee/information-holders/maksu-ja-tolliamet

[17] https://www.evs.ee/en/general-principles

[18] https://eelnoud.valitsus.ee/main#LwW8yZbl

[19] https://www.riigiteataja.ee/en/

was discovered that the information was formatted as tables in pdf documents, and would likely require data processing to be used in the analysis. The portal also did not offer API services to gather data automatically, making data extraction and updating a time-consuming task.[20]

REA: It is the Register of Economic Activities (*Majandustegevuse Register*) which provides information about licenses and notices related to some of the different areas of activities. Information on obligations relating to different sectors and businesses could be searched on the portal, as well as corresponding documents on notices of economic activity and activity licenses. The service allowed for searching, but did not directly provide any API service support to gather information automatically. For each area of activity, there was a list of contact persons whom to reach out to. The portal was linked to State Legislation Documents (*Riigi Teataja*), and documents related to obligations and acts could be downloaded as pdf files.[21]

SSSC (RTK): The State Support Services Centre (*Riigi Tugiteenuste Keskus*) belongs to the administrative area of the Ministry of Finance. The service prepares consolidated state accounting reporting and organised public procurement. Moreover, it distributes grants to companies in the following areas: entrepreneurship and innovation, labour market, education and science, culture and sports, integration, regional development, transport, social protection, and health. The portal provided an interface to do registration requests and to submit information requests. However, during the assessment, it was identified that there were no direct links on the portal where the data can be gathered immediately.[22]

ARIB (PRIA): The Agriculture Register and Information Board (*Põllumajanduse Registrite ja Informatsiooni Amet*) operates under the authority of the Ministry of Rural Affairs. It supports rural development in Estonia and all their activities which aimed to improve the quality of countryside life. The ARIB service portal provided an information centre where all obligations and fields of activities related to rural development and businesses could be found and downloaded in pdf format. However, no API service had been identified to gather information automatically .[23]

CEI (KIK): The Centre for Environmental Investments (*Keskkonnainvesteeringute Keskus*) is the implementer of modern, smart and effective solutions in the Estonian environmental landscape. The authority provided information on current grants for open projects, as well as on projects that have already been implemented, including all materials necessary for implementation. There was a data system available on the service portal that required pre-authentication before accessing the data. During the assessment, it was discovered that the portal did not offer an interface for data extraction and the quality of the data was also unknown and appeared to be in the form of unstructured text.[24]

Estonian Public Procurement Register (Riigihangete Register): It provides an environment for contracting authorities to conduct and economic operators to participate in public procurements. Enterprises can search published procurements, notices, contract information and review committee's decisions. The user guidelines were available on the service, as well as different search criteria for procurement-related information. The data related to public procurement in Estonia is available through an open data portal. It was possible to obtain information on procurements and contracts as published in relevant notices. The xml of procurement notices and contract award notices were in xsd 2.0.9 schema (defence sector notices in 2.0.8 schema). The data is renewed once per 24 hours at night. The data could be gathered through either manual pdf downloads or via API in a readable format. [25]

TED: Tender Electronic Daily provides free access to business opportunities from the European Union, the European Economic Area and beyond. The service compiles and publish data related to procurements on a daily basis. There was a search functionality allowing to find procurements of interest by different regions, countries and business sectors. Information about every procurement document is

[20] https://emtak.rik.ee/EMTAK/pages/klassifikaatorOtsing.jspx

[21] https://mtr.ttja.ee/?m=1

[22] https://www.rtk.ee/asutus-uudised-kontakt/kontaktid/dokumendiregister

[23] https://www.pria.ee/infokeskus

[24] https://kik.ee/et/toetatavad-tegevused

[25] https://riigihanked.riik.ee/rhr-web/#/open-data

published in the 24 official EU languages. All notices from the EU's institutions are published in full in these languages. The data is available through RSS feed and data can be extracted through it.[26]

e-ETCB training: Estonian Tax and Customs Board provides and organises free-of-charge training courses for taxpayers, and it requires electronic registration to participate. There are also European Commission e-courses and European Commission's Taxation and Customs Union e-courses available on different tax and customs topics, e.g. VAT refund, EMCS – Excise Movement and Control System, permit training etc. During the assessment, no direct data interfaces were identified that could be used to extract data automatically, making data availability related to information on tax training low.[27]

visiidid.ee: Visit Estonia provides the service related to state, official and working visits in partnership with the Ministry of Economic Affairs and Communication and Enterprise Estonia. The portal provided an interface to view and search for different courses and visit programmes planned in different business sectors, locations and times. The portal's main service is to help Estonian businesses to increase their visibility in external international markets. During the assessment, no direct data interfaces were identified that could be used to extract data automatically, making data availability on courses low.[28]

EAS: Enterprise Estonia provides services to contribute to achieving the long-term strategic goals of the Estonian economy. The platform aims to support the process of Estonia's transformation into one of the most competitive countries in the world. It provided different services to Estonian enterprises, from finding opportunities to getting funding support. The portal contained information related to grants, events and training but lacked data interfaces that could be used to gather and utilise this information. The data might be available if requested.[29]

Architecture assessment

Moreover, the technology architecture has been studied to assess the architecture requirements of AI-enabled services within the DGE environment. The DGE environment is based on current state portal system. Figure 5 below presents an identified existing architecture of the DGE portal. The AI environment and components are to be added to the target architecture which is discussed in chapter 5.

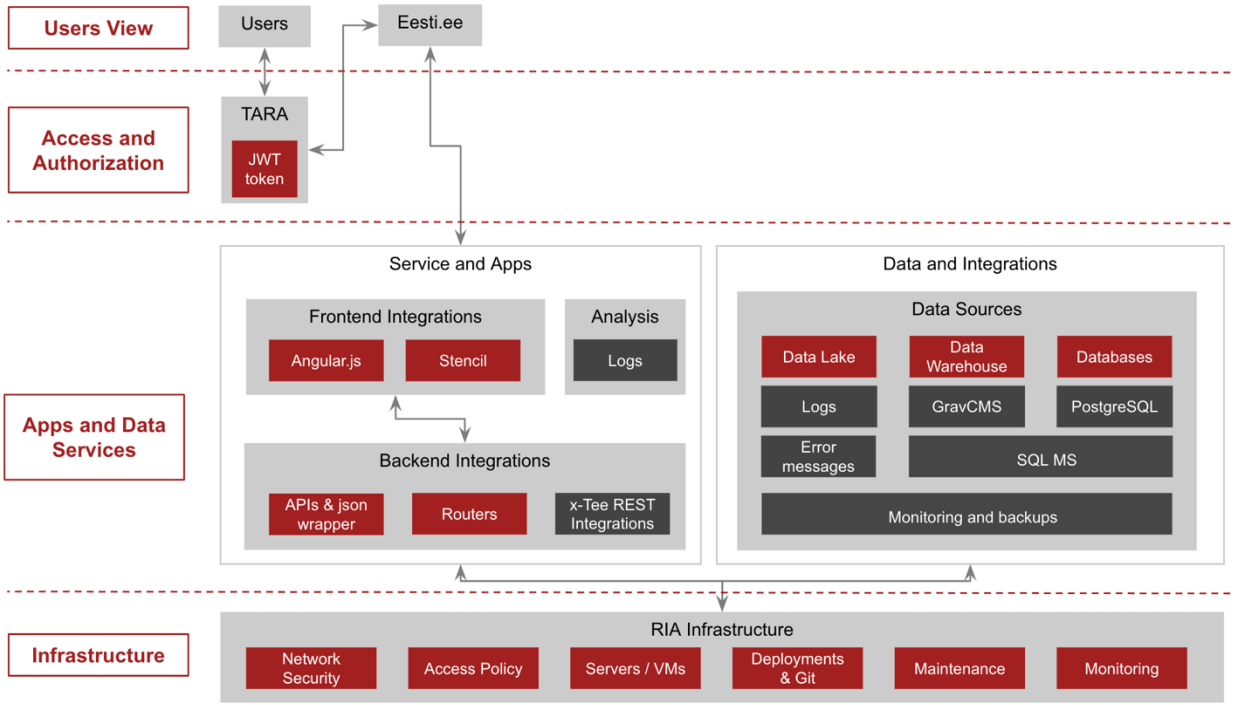
[26] https://ted.europa.eu/TED/main/HomePage.do

[27] https://www.emta.ee/en/business-client/e-services-training-courses/training-courses

[28] https://visiidid.ee/

[29] https://eas.ee/en/

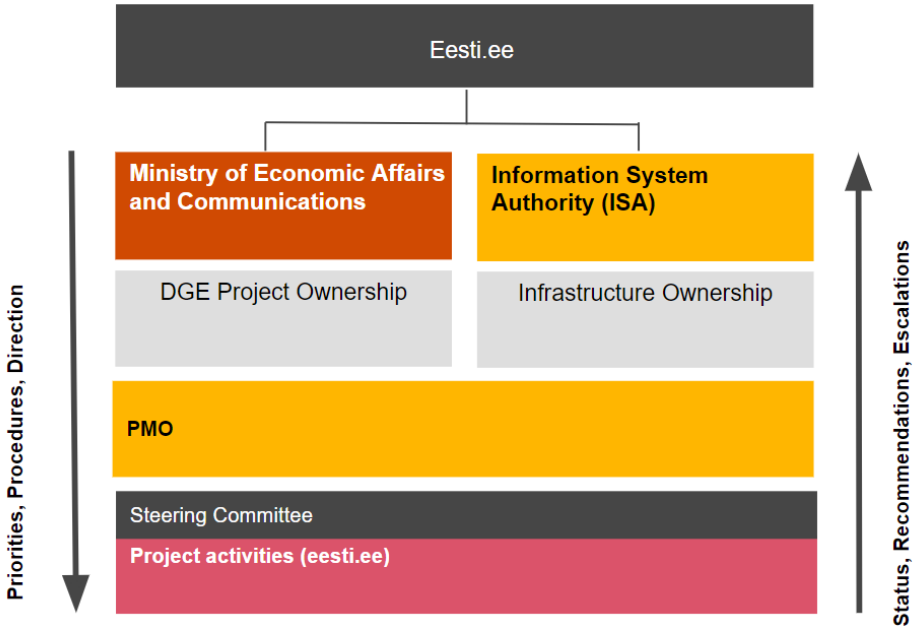
Figure 5. Overview of current DGE architecture



Governance and process assessment

Furthermore, Figure 6 models the current governance of the state portal which would be transformed into a new structure along with change management and AI development. Below is a high-level definition of the existing governance structure of the state portal.

Figure 6. High-level current governance structure



3. Selection criteria for DGE AI use cases

Defining the criteria to evaluate and select the most beneficial AI use cases was the critical step which was completed within different workshops, and feedback sessions with the beneficiary. The aim was to create criteria that cover all the important aspects in a simplified way. During the discussions, it was identified that the following criteria covered the aspects well and were suitable for evaluating DGE use cases.

This chapter provides an explanation of the selection criteria for evaluating AI use cases. It describes in detail the underlying dimensions for each of the criteria and the methodologies defined to evaluate and score them.

3.1 Methodology of selection criteria measures

The "impact" and "ease of implementation" measures were used to define the selection criteria. These measures helped assess and evaluate the "cost-benefit" for a particular AI use case. The impact measured the overall benefit and impact of the use case, whereas "ease of implementation" attempted to assess the complexity of implementation and cost aspect.

3.1.1 Measure estimated impact

The measures used to estimate the impact of the use cases were the following, especially three measures: “expected benefit”, “no. of entrepreneurs affected” and “expected frequency of use”. These measures provided a good baseline for estimating the impact and importance of the use case solution considering DGE's purpose.

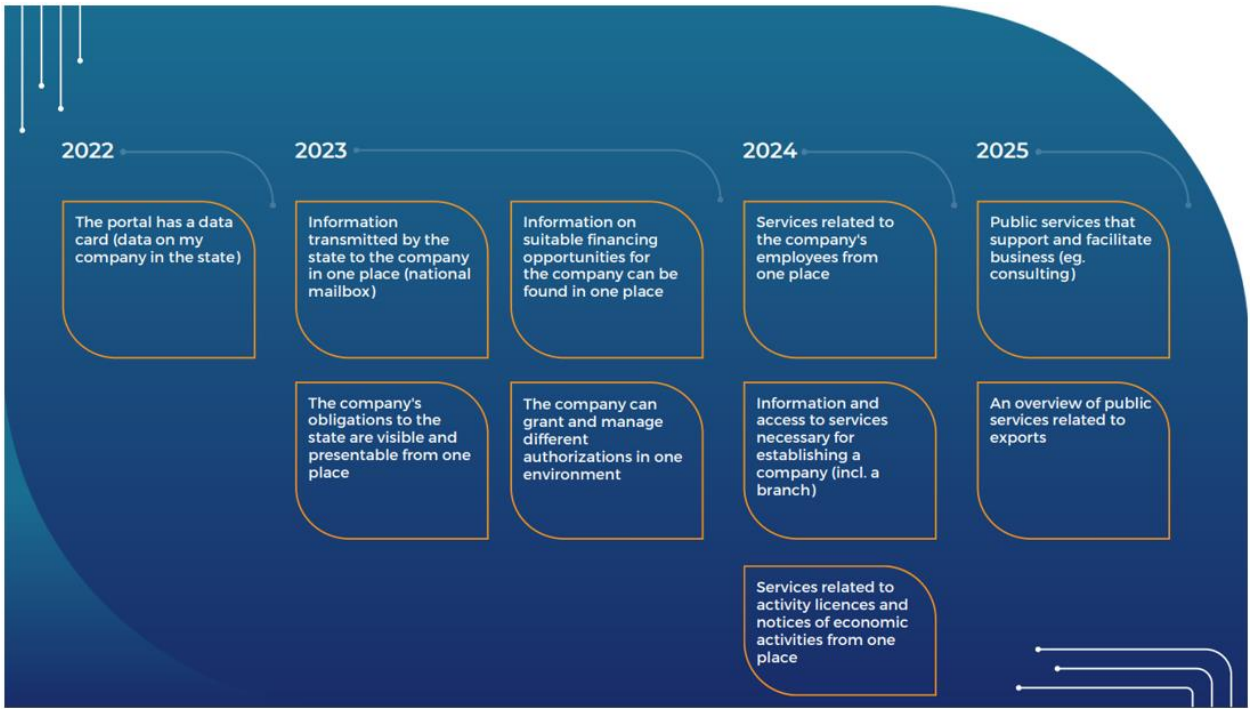
Expected scalability

Scalability was defined as the ability of a solution to accommodate an increased workload. It indicated the ability of a solution to handle repeatedly extensions to the solution's capacity.

In our scenarios, an "expected scalability" of a use case was the ability to accommodate other use cases. Moreover, it considered the top ten of DGE's IT-development projects. Figure 7 presents the top ten DGE’s IT projects.

The 1-3 scales were used to evaluate the "expected scalability". The order of measurement was in ascending order, e.g. a low score of 1 indicated that the use case had low scalability, and a high score of 3 showed that the use case had high scalability to other use cases as well as to the top ten DGE's IT-development projects.

Figure 7. Overview of the ten IT development projects planned for the DGE



Expected benefit

The expected benefit was a critical measure in assessing the impact of the use cases. It was identified that it should be covered and especially from the aspect of how the specific use case could benefit entrepreneurs and enterprises.

The user needs gathered during the D3 were used as a baseline to measure this sub-criterion. The user had to describe the areas where entrepreneurs needed support and efficient and effective services from the DGE. Therefore, the hypothesis was developed that if the AI use case addressed most of the user's needs, it should be the most beneficial one for enhancing the DGE's service. The hypothesis was discussed and communicated to the beneficiary and stakeholders before measuring the "expected benefit" of a use case.

The scale of 1-3 was used to evaluate the "expected benefit ", in an ascending order. For example, a low score of 1 indicated that the use case had low benefit, and a high score of 3 showed that the use case had high benefit.

Number of entrepreneurs affected

During the different sessions with the beneficiary, it was identified that the "expected benefit" measure only considered the user needs and did not incorporate the number of entrepreneurs. Since some of the user needs gathered during the D3 only depicted the needs of a small sample of entrepreneurs interviewed during the workshops, there had to be a measure that assessed the "number of entrepreneurs affected" by an AI use case.

Simple cut-off percentages were used as baselines to measure this sub-criterion. For example, if a use case was assessed to affect 20% or fewer entrepreneurs then the "no. of entrepreneurs affected" by a use case was low. Similarly, if use case was estimated to affect 50-60% of entrepreneurs, it was considered medium, and over 60% was considered high. The hypothesis was discussed and communicated to the beneficiary and stakeholders before measuring the " no. of entrepreneurs affected " by a use case.

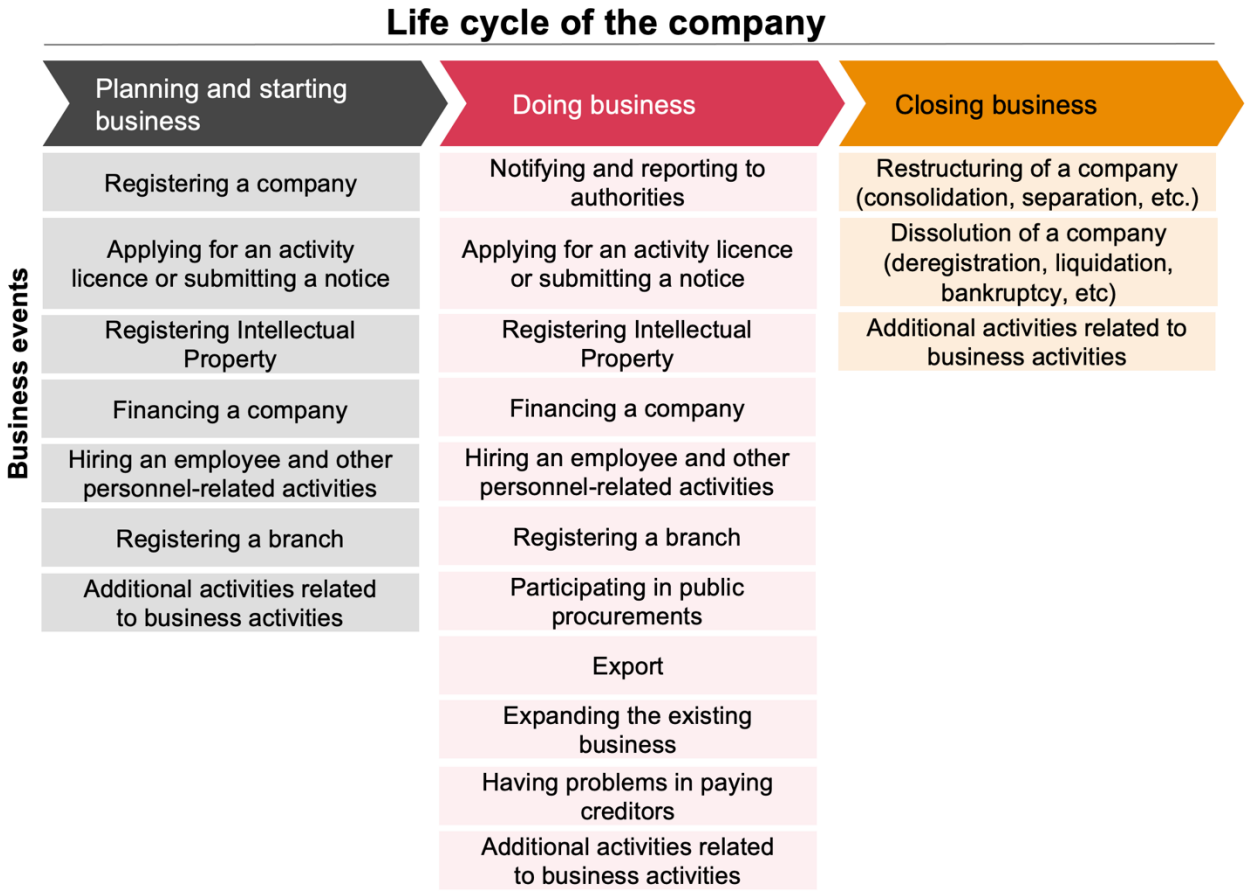
Expected frequency of use

Furthermore, during the workshops and sessions, it was highlighted that business events were also important for measuring the impact of the services. For example, some business events only occur once in an enterprise's life cycle, e.g. establishing or dissolving a company. Similarly, some business events related to tax submissions, financing and business operations occur frequently.

Therefore, the sub-criterion "expected frequency of use" was chosen as part of the measurement of the overall impact of a use case. The purpose was to assess use cases that solved frequent problems encountered by entrepreneurs.

Figure 8 describes the three phases of a company's life cycle and the business events corresponding to them. The business events described in Figure 8 were used as baseline events for assessing this sub-criterion.

Figure 8. Life cycle of the company with main business events considered in DGE projects.



The scoring was based on a scale of 1 to 3 in ascending order. For example, a high score of 3 indicated that the use case addressed a business event that occurs frequently, at least once a month, whereas a low score of 1 indicated that the use case focused on business events that occur less frequently, such as once a year or less.

3.1.2 Measure estimated ease of implementation

The measures used to estimate the ease of implementation for the use cases were the following:

Data readiness

During the workshops and sessions, it was identified that data readiness was one of the most critical factors in assessing the implementation aspect. Supporting data was essential to build an AI use case, and if there was no data available or if the quality of the data was poor, the implementation of a use case would become more difficult.

To assess the data readiness, four underlying data aspects were considered:

1. Data availability
2. Data quality
3. Data volume
4. Data preparation

The data preparation factor was added because it was important to consider from an AI modelling perspective. For example, if the data was unstructured, more effort would be required to prepare and label the datasets before AI modelling.

A scale of 1 to 3 was used to measure the score for this sub-criterion. A high score of 3 indicated that the data was of good quality and readily available in large volume and could be easily prepared to train AI models for a specific AI use case.

Lastly, to cover unknown scenarios in the evaluation, it was decided during the sessions that if the information related to data sources was limited, a score of 1 should be assigned. It would indicate that the data readiness was low.

Technology and infrastructure

Technology and infrastructure were added as another sub-criterion to measure the ease of implementation of a use case. During the workshops and sessions, it was discussed and communicated with the beneficiary that the following underlying aspects would be considered to score this measure:

1. AI algorithms complexity
2. Compute resources to train, test and consume AI models
3. Network infrastructure
4. Storage capacity and security

This sub-criterion was scored on a scale of 1 to 3, where a high score of 3 indicated that the use case did not require complex technology and infrastructure. For example, the best-case scenario would be that non-complex algorithms were used for AI models and there was no need to use expensive computing resources.

Moreover, network infrastructure and security aspects were also covered in the assessment because in some scenarios there was a need for high-bandwidth and low-latency networks. However, during the assessment, it was identified that the DGE AI use cases did not have such corner scenarios.

Development cost

The cost of development as a factor to assess the ease of implementation was also added to the cost-benefit assessment. During the assessment, it was identified that cost estimation should be made for the MVP-level solution. Therefore, the development cost of implementing an MVP for a specific use case was considered when evaluating the use cases.

A scale of 1 to 3 was used to score this measure. However, in this scenario, a high score indicated a low score of 1 because if the cost of development were high, the ease of implementation would be difficult. The number of estimated development hours was used as a baseline for assessing and scoring this measure. For example, if the estimated number of development hours for an MVP of a use case was 450 hours or less, the development cost would be scored as low.

3.2 Methodology of selection criteria scales

3.2.1 Method to score AI use cases

Table 2 and Table 3 below provide descriptions and scales for our selection criteria's measures.

Table 2. Overview of the impact criteria measures and scale

Impact				
Criteria	Low	Medium	High	Scale
• Expected scalability	• Use case that had limited scalability to other use cases and DGE's IT projects.	• Use case that could be scaled to 2-3 other use cases or to DGE's IT projects.	• Use case that could be scaled to 4+ other use cases or DGE's IT projects.	low = 1 medium = 2 high = 3
• Expected benefit	• Use case that addressed only a single user need identified in deliverable 3.	• Use case that addressed 2-3 of the user needs identified in deliverable 3.	• Use case that addressed 4+ of the user needs identified in deliverable 3.	low = 1 medium = 2 high = 3
• No. of entrepreneurs affected	• Use case that would potentially impact 20% or fewer entrepreneurs.	• Use case that would potentially impact 50-60% of entrepreneurs.	• Use case that would potentially impact 100% of entrepreneurs	low = 1 medium = 2 high = 3
• Expected frequency of use	• Use case that impacted infrequent business events that occur only once a year or less often.	• Use case that impacted business events that occur once a quarter.	• Use case that impacted business events that occur frequently, at least once a month or more often.	low = 1 medium = 2 high = 3

Table 3. Overview of the ease of implementation criteria measures and scale

Ease of implementation				
Criteria	Low	Medium	High	Scale
• Data readiness	• Use case that had uncertain access to data. Also, the data might be distributed among several sources and there was low confidence in its quality. Moreover, the data was in an unstructured format.	• Use case where data was partially available. Also, it might be distributed among several sources with a large volume and a high degree of confidence in its quality. Moreover, a considerable amount of data was in a semi-structured format.	• Use case that had data available in one major source system, with a large volume and a high degree of confidence in its quality. Also, the data was in a semi-structured or structured format.	low = 1 medium = 2 high = 3
• Techno-logy and	• Use case that had a high degree of	• Use case that had a medium degree	• Use case that had a low degree of	high = 3 medium = 2

Ease of implementation				
Criteria	Low	Medium	High	Scale
infra-structure	complexity considering the technology and infrastructure. For example, AI algorithms of high complexity, high computing resources (GPUs) and high level of storage requirements.	of complexity considering the technology and infrastructure. For example, less-complex AI algorithms. Medium computing (GPUs) and storage requirements.	complexity considering technology and infrastructure such as non-complex AI algorithms, fewer computing resources and storage requirements.	low = 1
• Develop-ment cost	• Use case where an MVP implementation was considered a large project, e.g., 600+ development working hours.	• Use case where an MVP implementation was considered as a medium project, e.g., 500-600 development working hours.	• Use case where an MVP implementation was considered as a small project, e.g., less than 450 development working hours.	low = 3 medium = 2 high = 1 (This value was reversed from this scale to evaluate its impact on ease of implementation)

4. Preparation and evaluation of the DGE AI use cases

The use cases were prepared based on the user needs gathered in Deliverable 3 and evaluated based on the selection criteria defined in chapter 3. This chapter provides detailed informatio1n on the identified AI use cases and how these use cases were evaluated and scored using the selection criteria. Moreover, it presents detailed use case cards for the two selected use cases that were selected after the evaluation and validation process.

4.1 Prepare a list of AI use cases

The DGE AI use cases were identified and prepared in various iterations. Different workshops and sessions were held with the beneficiary to discuss and present the use cases. Initially, a list of more than twenty-five AI use cases was identified that corresponded to the user needs and different AI-oriented ideas.

The initial list of AI use cases was filtered down to include only those that were aligned to the DGE context. The final list of use cases, shown in Table 4 below, was shortlisted to meet the user needs identified in D3.

Table 4. List of AI use cases derived from user needs

#	Use case	Description
1	Increase DGE user engagement with personalised recommendations using AI	This use case intended to enhance personalisation based on a user's recent activity on the portal. The aim was to develop an AI solution that used session-based recommendations that were dynamically adjusted based on the user's recent activities on the state portal.
2	Enhancing the process of providing license and compliance information to entrepreneurs in the DGE using AI	This use case intended to enhance the DGE channel to provide users with the most relevant content, especially compliance and licence-related content depending on the enterprises' field of activity. Each field of activity and its divisions often contained specific compliance requirements e.g., forestry, mining, manufacturing. Providing such compliance information to business effectively could be beneficial. The aim was to develop an AI-based solution that could take input related to the nature of business and recommend information to entrepreneurs for the next steps, e.g., relevant activity and its divisions, listing the relevant compliance clauses and licenses required.
3	Improve the process of providing tax and reporting information to entrepreneurs in the DGE using AI	This use case intended to improve the process of providing the most relevant content to users, especially related to taxes and reporting or other relevant factors, based on the enterprises' field of activity and other related business events. The aim was to develop a solution that could predict and recommend information to entrepreneurs for the next steps, e.g., event-driven reporting.
4	Improve the DGE information channel for entrepreneurs to share and suggest business-related information using AI	This use case aimed to apply AI to solve the problem of finding relevant information on a wide range of business-related topics. The idea was to use advanced topic modelling techniques on unstructured data related to tax, obligations, compliance, regulation, etc. Entrepreneurs could subscribe to different topics and receive relevant information.

#	Use case	Description
5	Guide entrepreneurs through the DGE's smart suggestions channel for deadlines and obligations	Enterprises are constantly working to meet deadlines and obligations. This use case aimed to use an AI solution to enhance the process of suggesting to entrepreneurs how they could move forward with their obligations and specific deadlines.
6	Recommendations for training and events in the public sector	This use case aimed to enhance the process of sharing public sector-related events (e.g., conferences, training) and information for entrepreneurs using AI. Text mining AI could be applied to data on companies, entrepreneurs and business activities and then suggest relevant public sector events or articles.
7	Improve suggestions related to public sector financing opportunities	This use case focused on improving the process of providing companies and entrepreneurs with information on public sector funding opportunities using AI. It covered information related to co-financing, grants, loans, innovation funds, guarantees, and other related support offered by public services. Information related to procurement opportunities would not be considered in this use case. Furthermore, this use case might take into account the information related to the financial condition of enterprises, if available or shared by consent. Such information would enhance AI-based suggestions for enterprises.
8	Smart search for EU financing opportunities	This use case aimed to improve the process of sharing information on direct EU financing opportunities for Estonian entrepreneurs and enterprises. The focus was on extracting information related to grants, funds, loans and guarantees offered by the EU and making it available to Estonian companies to take advantage of these financing opportunities. Processing information related to procurement at the EU level is beyond the scope of this use case.
9	Public procurement recommendations	This use case aimed to assist enterprises in the processes related to public procurement and tender submissions using AI. A solution could use previous submission history to provide enterprises with recommendations on how to apply and participate in different public procurements.
10	Improve the process of finding the right experts in the public sector	This use case aimed to improve the process of finding the right experts in the public sector through the DGE channel to solve complex business-related issues. Entrepreneurs could share what they are looking for and AI could be used to search through unstructured information and suggest the relevant experts and resources.
11	Smart DGE search engine	This use case aimed to address the application of smart AI-based search in the DGE channel. AI could be used to create a knowledge management system which possessed all levels of details related to the fields of activity and events of the business life cycle in Estonia. Entrepreneurs could use this search to quickly search for any topic of interest.
12	Smart suggestions to improve business export and growth	This use case aimed to provide support to enterprises that export goods and services. The focus was to help enterprises through smart analysis to improve their current productivity and market share, as well as to expand their services. Smart suggestions could be related to, for example, highlighting and suggesting that their export-related business

#	Use case	Description
		had been growing and there was generally room for increased profits through innovation and business expansion. Moreover, providing on a high level the information related to required regulations and obligations. The suggestions would not be linked to the search for grants or loans, as there were other use cases and existing services that offered such support to entrepreneurs.

4.2 Prepare data sources list for AI use cases

The most critical step in assessing and evaluating the use cases was to identify the essential data elements. The data elements set the base of any AI solution and if data is not available or ready to be used, the solution cannot fulfil the AI objectives.

4.2.1 List of potential data required

The list of data elements was prepared considering the identified use cases. It showed at a high level what data and information should be assessed to evaluate the use cases. The assessment of data readiness particularly depended on the assessment and analysis of these data elements.

Chapter 2 explains in detail how these data elements were analysed and which data sources could be used to collect data. Table 5 below provides a list of data elements that might potentially be required to assess and evaluate the AI use cases.

Table 5. List of the data elements required to assess and evaluate the use cases.

Data elements	Description
User/entrepreneur information	Basic profile details Activity performed on the state portal such as clicks, views and submissions.
Enterprise data	Detailed company data. Company's historical events/submissions data. Company's licenses, compliance information and notices of economic activity. Company's field of activities information. Company's public data related to taxes and reporting
State portal statistics	User sessions {session length, action taken, timestamps} Content and articles items and information. Clicks, views and submission-related information
Licenses and notices information	Licenses and notices requirements data and information for different fields of activities, both structured and unstructured
State portal statistics for licenses and notices sections	Information related to views and clicks related to licenses and notices sections.
Tax and reporting information	Tax and reporting requirements related information for different fields of activities, both structured and unstructured
State portal statistics for tax and reporting sections	Information related to views and clicks especially for sections related to tax and reporting
Information on fields of activities and business events	Documents and content related to different fields of activities and business events
Training and conferences information	Tax and custom training information and other training and conferences content for different fields of activities

Data elements	Description
Financing and grants information across Estonia and the EU	Public sector grants and funding information and other financing opportunities content for different fields of business and enterprises. Also, the historical training received in a peer group
Public procurement information	Information on new tenders, historical data related to old submissions, information, and content relevant to applications, enterprises' history of submission for tenders
Information of public sector experts and FAQs	Experts' information in different fields of activities, FAQs information from different portals related to business events & fields of activities and other relevant information.

4.3 Evaluation of DGE AI use cases

The evaluation of AI use cases was completed in several sub-phases and during communication with the beneficiary. First, the PwC team conducted the first assessment and assigned scores to each of the selection criteria. As explained in chapter 3, each criterion had a scale of 1-3 points and for each use case, the scores were initially assigned by the PwC team.

Moreover, during the evaluation process, descriptions of use cases and selection criteria were enhanced. Also, different sessions with the beneficiary brought more information which helped during throughout the evaluation process.

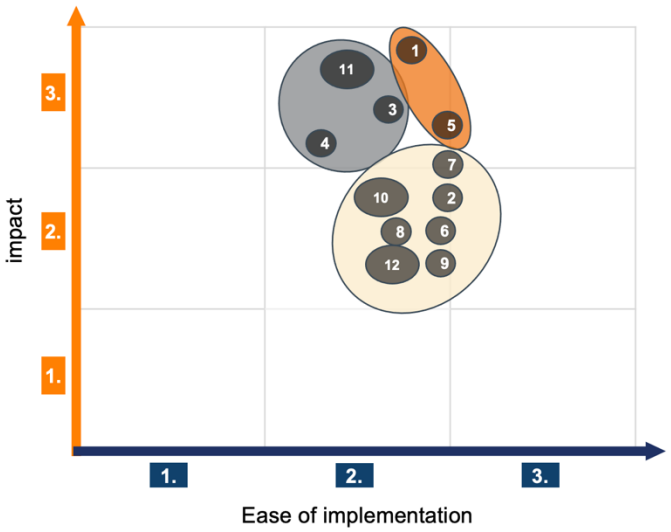
4.3.1 Scores of DGE AI use cases

The key findings of the use case assessment were finalised during the communication with the beneficiary and are presented in the following Figure 9. It shows the different scores for each of the use cases in terms of "impact" and "ease of implementation" criteria. Moreover, the detailed excel-based evaluation is described in Annex 1. During the evaluation, it was found that regarding ease of implementation use cases were not very differentiated.

Use case #1 "increase DGE user engagement with personalised recommendations using AI" received the highest score for impact and benefits, while use case #5 "guide entrepreneurs through the DGE's smart suggestions channel for deadlines and obligations" received the highest score for "ease of implementation".

At a high level, three clusters were identified for the use cases based on the similarity in evaluation scores, as shown in Figure 9 below. Use cases 1 and 5 in the "orange" cluster are presented as the most beneficial use cases in terms of impact and ease of implementation point of view.

Figure 9. Evaluation and scores of the use cases



#	List of DGE AI use cases
01	Increase DGE user engagement with personalised recommendations using AI
02	Enhancing the process of providing license and compliance information to entrepreneurs in the DGE using AI
03	Improve the process of providing tax and reporting information to entrepreneurs in the DGE using AI
04	Improve the DGE information channel for entrepreneurs to share and suggest business-related information using AI
05	Guide entrepreneurs through the DGE's smart suggestions channel for deadlines and obligations
06	Recommendations for training and events in the public sector
07	Improve suggestions related to public sector financing opportunities
08	Smart search for EU financing opportunities
09	Public procurement recommendations
10	Improve the process of finding the right experts in the public sector
11	Smart DGE search engine
12	Smart suggestions to improve business export and growth

4.4 Validation of the evaluation results

To validate the evaluation results a survey was conducted among Estonian entrepreneurs. The survey was created using Google Forms and sent directly by email to approx. 2,000 SMEs between 21st of February and 1st of March. The contact information was requested from the national e-Business Register. The survey was in Estonian, and the AI use cases were presented as potential solutions in a simplified form (see Annex 2). Each use case was accompanied by a practical example.

The participants were asked to rate each use case on a scale from 1 to 5:

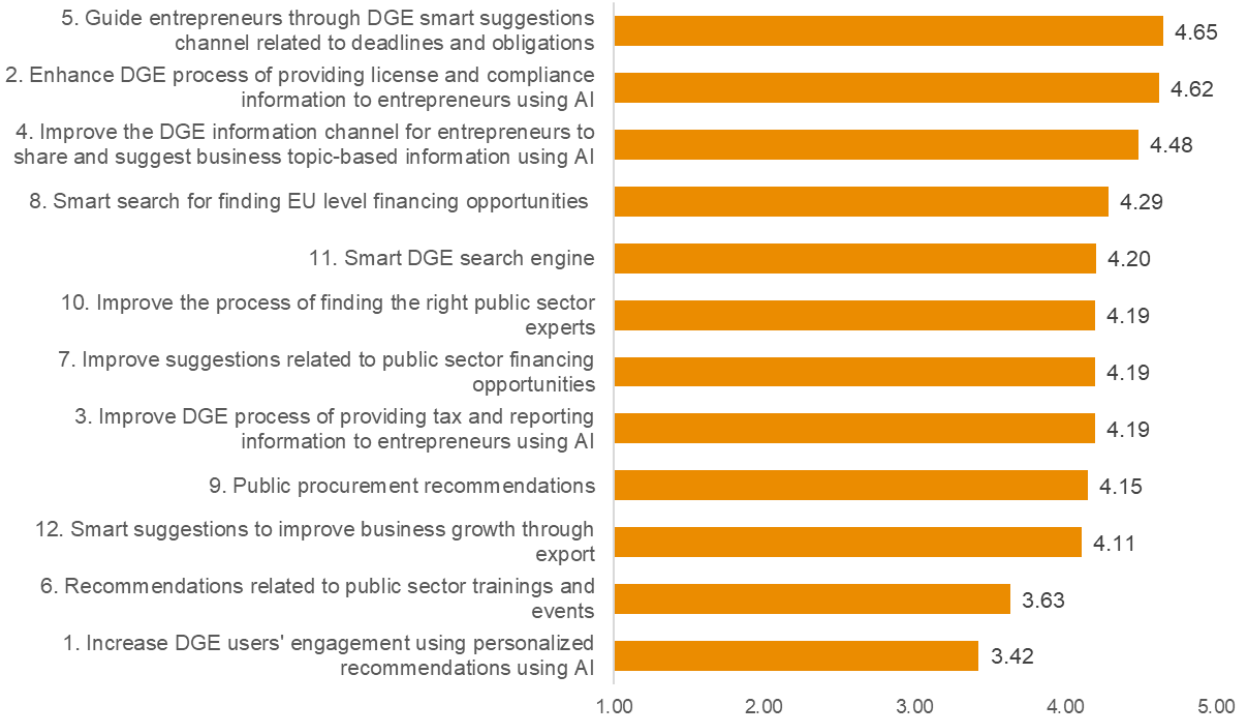
- Not useful at all
- Rather not useful
- Neutral
- Rather useful
- Very useful

The participants also had an option of marking the solution as "not relevant" if it was not suitable for them.

31 responses were received, meaning that approx. 1.6% of the sample responded. Among the respondents, 62.5% were companies with fewer than ten employees, 18.8% had zero employees, 12% had fewer than 50 employees and 6.3% had fewer than 250 employees. The participants were mainly from the manufacturing, wholesale, and retail, information, and communication sectors.

Participants gave the highest scores to use cases number 5, 2 and 4, as presented in Figure 10.

Figure 10. Average scores of the perceived usefulness by entrepreneurs



Since the number of responses was relatively small, it was decided to further validate the results with the help of representatives of the Estonian Chamber of Commerce and Industry (ECCI) and the Estonian Association of SMEs (EASME). Two separate interviews were organised on 2nd of March and 9th of March. Prior to the interviews, the same survey was sent to both representatives so they could also try to evaluate the use cases in the same manner. The results of the survey were presented during the interviews. We asked the representatives to comment on the entrepreneurs' scores and to present their own scores with explanations. The main comments and suggestions for each use case were as follows:

1. Increase DGE user engagement with personalised recommendations using AI

ECCI: To benefit from the solution, the user must make inquiries in the DGE. (Score 3)

EASME: The solution could be built to take into account the different roles the specific representatives have using the DGE environment (e.g. accountants, HR specialists, etc.). (Score 4)

2. Enhancing the process of providing license and compliance information to entrepreneurs in the DGE using AI

ECCI: These recommendations should include additional information on permits and notices so that the user can assess whether they need them. At the same time, the user should not be overloaded with information. (Score 5)

EASME: All information on licensing and economic notices is available in the Register of Economic activities. We should not create a solution where entrepreneurs become overly dependent on AI. It should be further investigated in which areas entrepreneurs have been late with their license applications. (Score 4.5)

3. Improve the process of providing tax and reporting information to entrepreneurs in the DGE using AI

ECCI: The user could have the option to decline the recommendations if they are already familiar with the process. (Score 4)

EASME: This solution is probably more focused on the role of the CFO and/or accountant in the company. Many small companies have outsourced their accounting services, so this solution is likely to have little impact on entrepreneurs. Accountants usually already have a good knowledge of the law and obligations. For a start-up entrepreneur, this could be more beneficial. (Score 3)

4. Improve the DGE information channel for entrepreneurs to share and suggest business-related information using AI

ECCI: The field of activity should be the main factor that sets the right direction. (Score 4)

EASME: Entrepreneurs could have the opportunity to choose the topics that are most relevant to them. It also depends on the different roles within the company. (Score 4)

5. Guide entrepreneurs through the DGE's smart suggestions channel for deadlines and obligations

ECCI: It is important to avoid double notifications – for example, if one notification comes from the Tax and Customs Board and another from the DGE. There should be an option for the entrepreneur to define which notifications should go to different roles (e.g., accountant, HR, etc.). (Score 4)

EASME: This should be well segmented to better understand different obligations (e.g., reporting) set by different state organisations. (Score 4.65)

6. Recommendations for training and events in the public sector

ECCI: It should be carefully considered whose training and seminars will be presented in the DGE, e.g., should it be limited to public sector training and events? Risk of creating unfair competitive advantages for private market players that provide training and seminars to entrepreneurs. (Score 2)

EASME: It should be considered whether the training and seminars are intended for entrepreneurs only or also for their employees. The quality of training varies greatly, what happens if entrepreneurs are not satisfied with the training? Instead of recommending training and seminars, the entrepreneurs could search for relevant training and seminars on the DGE platform using keywords. It was recommended to start with public seminars only. (Score 4)

7. Improve suggestions related to public sector financing opportunities

ECCI: Particularly useful if it considers the company's profile. (Score 5)

EASME: Local funding opportunities should be also considered. (Score 5)

8. Smart search for EU financing opportunities

ECCI: For some enterprises, it would be very useful, but most of them would probably not use it. (Score 2)

EASME: It would be great to have an overview of EU funding opportunities, but it is not very practical. Estonian entrepreneurs find it difficult to apply for EU grants. (Score 2.5)

9. Public procurement recommendations

ECCI: It might be irrelevant for most enterprises. Entrepreneurs should have the option to disable the recommendations. (Score 4)

EASME: There is already an existing solution (Estonian Public Procurement Register), where enterprises can search for procurements in areas of interest. It is risky to start making these recommendations within the DGE – what happens if in some cases entrepreneurs are not notified – who will be responsible for this? (Score 3)

10. Improve the process of finding the right experts in the public sector

ECCI: With complicated inquiries, it is probably hard to propose a specific organisation and a contact person. In general, it would be beneficial for entrepreneurs. This solution could also include contacts from different umbrella organisations. (Score 4)

EASME: For starters, a list of organisations with their web pages and general email addresses should be sufficient. (Score 4)

11. Smart DGE search engine

ECCI: Reasonable and beneficial solution. (Score 4)

EASME: The search engine is a must-have element of the DGE. Almost every system nowadays has a search engine that can provide relevant results. (Score 4.2)

12. Smart suggestions to improve business growth through export

ECCI: It is useful for entrepreneurs only if it can give very specific and accurate recommendations. It is a complex solution that would require a lot of resources. (Score 5)

EASME: It is designed for a small group of entrepreneurs only. This solution could provide an overview of the main export markets and how to enter those markets. For starters, DGE could provide information on export-related seminars, statistical overviews, grants and contacts. (Score 4.5)

Both representatives emphasised that the gateway should not overwhelm entrepreneurs with the information flow. If an entrepreneur starts receiving a lot of recommendations and notifications, they might become irritated and start avoiding them, therefore they might miss some very important notifications. Another risk is that they might become too dependent on them. Representatives also emphasised that to provide accurate recommendations, it is necessary to define the profile of the entrepreneur, i.e. what features should be considered when making recommendations (e.g. field of activity, age, size, role, etc.)?

4.5 Selected DGE AI use cases

This sub-section provides information for the selected use cases. After the evaluation, survey and interviews, the following two use cases were selected for the implementation model and roadmaps. Figure 11 and Figure 12 show the detailed cards for the selected use case #1 and use case #5.

The details describe the high-level information on important aspects related to selected use cases. For example, description, outcome, user needs addressed by the use case, potential data sources required, tools and technology considerations. Furthermore, Chapter 6 outlines detailed roadmaps for each of these selected use cases.

Figure 11. Details card for selected use case #1

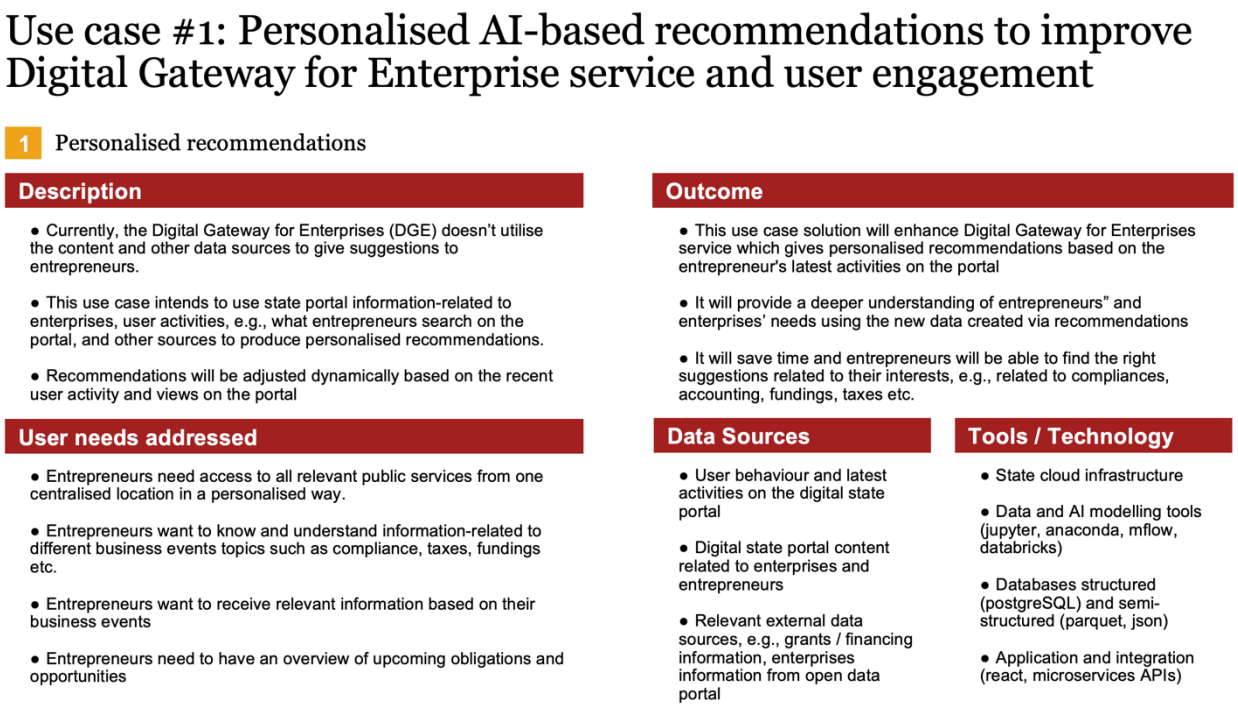
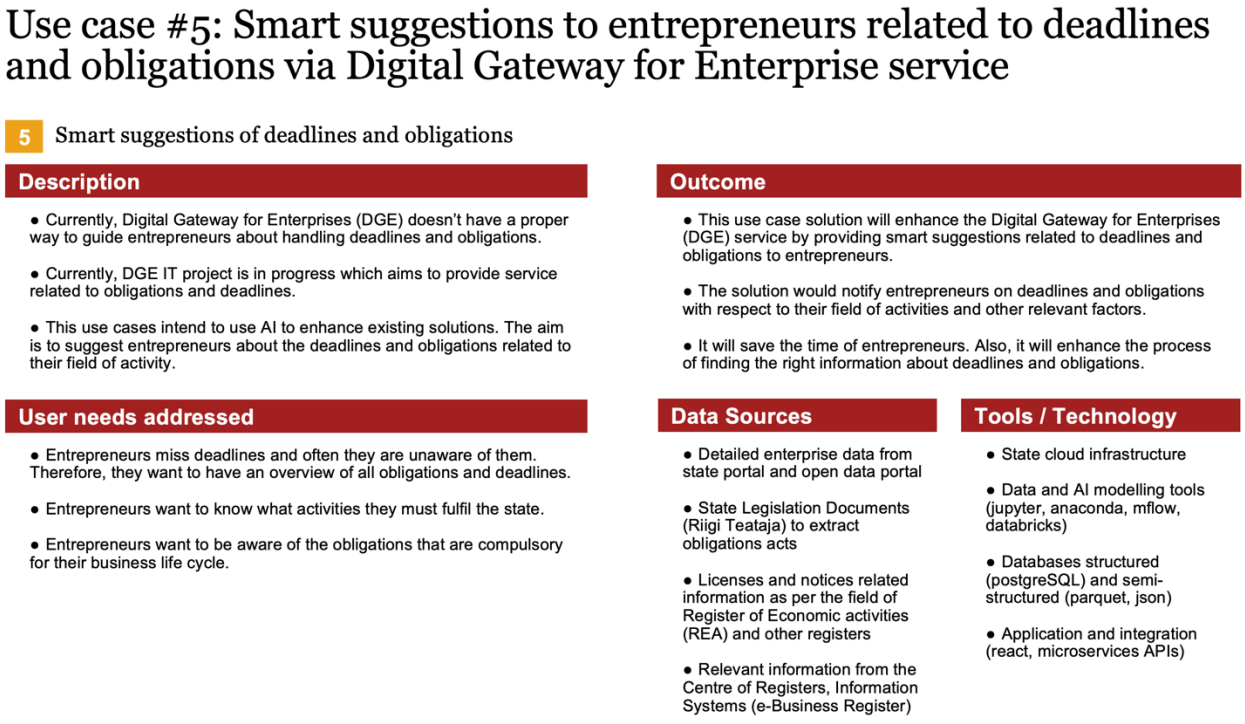


Figure 12. Details card for selected use case #5



5. Implementation of AI in the digital gateway for enterprises

Understanding the impact of infrastructure is inherently complex and has its limits. Digital services and the infrastructure that enable them must be understood as offering new choices and new capabilities.

Estonia consistently ranks as one of the most digital countries and has become the digital powerhouse, integrating technologies like digital ID, blockchain-enabled voting, and decentralised data storage into its government's day-to-day operations.

With the highest number of unicorns per capita, Estonia ranks high among the list of most open economies in the world as well. The digital gateway of the enterprise's infrastructure offers solutions to businesses — digital signatures, electronic tax claims and the availability of public records online — to provide transparency and efficiency. Particularly, Estonia's state portal and e-Business Register are advanced services that allow entrepreneurs to register and operate businesses online.

The overall landscape of public sector services which target entrepreneurs is still fragmented across several platforms. However, Estonia's state portal is bringing a unified digital experience to entrepreneurs, which can further be advanced using AI. Also, there is a potential to utilise AI to bring enhancements as well considering other public institutes and services.

The selected AI use cases can advance the services of digital gateway of enterprises but require solid AI-enabled digital infrastructure to support the AI-based solutions and services. During our workshops and interviews, it was assessed that the supporting infrastructure exists, however, the quality of AI service provision is low. This is due to several factors, among which data and organisational readiness for AI are critical.

This chapter provides an overview of how to enhance the existing digital gateway of an enterprise's infrastructure to leverage data and AI capabilities. Moreover, it would show where to focus to ensure that AI-based services provide a seamless, customer-centric, proactive and reliable experience for entrepreneurs.

5.1 Methodology framework: PwC Analytics & AI transformation framework

We demonstrate the state for the implementation of AI in DGE's infrastructure using **PwC Analytics & AI transformation framework,** which consists of the following six categories:

- Technology and Infrastructure (existing infrastructure, data storage)
- Data and Information (data structure and quality, data accessibility, data management)
- Legal (cybersecurity, data protection)
- Process and Integration (AI development and implementation, procurement process, AI integration with the existing system)
- Organisation, Culture and Talent (organisational structure, strategy, knowledge and competencies, roles, and responsibility)
- User-centric design (recognising users, mapping user needs).

5.1.1 Technology and infrastructure

The technology and infrastructure to enable and sustain AI services requires separate capabilities and resources. The entire AI development lifecycle, from experimentation to production deployment, needs a unified and secure technical infrastructure.

For example, during our workshops, several AI use cases were identified, and some of them were selected to enhance the digital gateway for enterprise services. Implementing those use cases, and then gradually moving towards data-driven and AI-driven services, required secure AI sandbox environment, fast and error-free integration of AI-related libraries, data lake-like capabilities to store (cross-use and exchange data) experiments, and servers with different computing capacities.

Technical infrastructure and a proper channel to provide resources and services would be the necessary step to accelerate the development of AI in the digital gateway to enterprise infrastructure. In an increasingly AI-influenced world, as AI moves from experimentation to adoption, it demands significant computing resources and technical infrastructure.

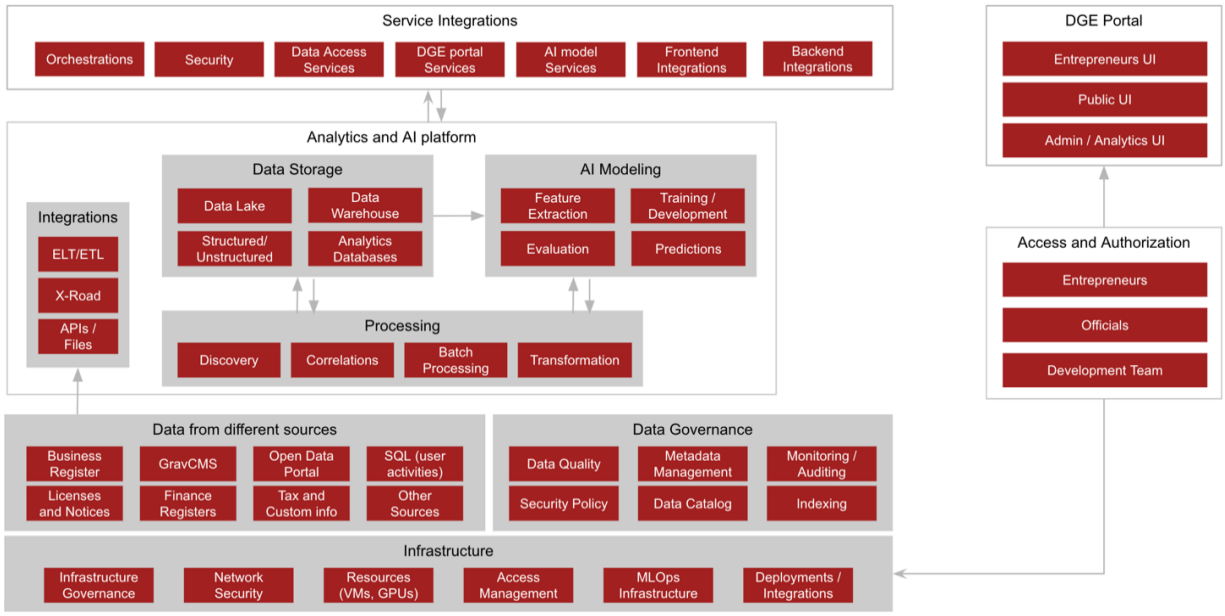
During our interviews and based on the preliminary AS-IS analysis, it was identified that the state portal uses the state cloud (Riigipilv) as technical infrastructure support. It would be useful to highlight relevant information related to current and future planned improvements to the state cloud (Riigipilv). The state cloud (Riigipilv) has faced a 57% increase in resource utilisation since February 2022, which has led to a greater load on the infrastructure, but the infrastructure expansion has been hampered by long delivery times. Rising electricity prices have led to an increase in infrastructure costs, but the cost of resource use for clients has not changed.

The next state cloud (Riigipilv) 1.1 includes infrastructure expansion with new servers and network devices, network harmonisation, the introduction of SDN (Software Driven Network), the expansion of supported platforms and the modernisation of the self-service platform.

The future of state cloud (Riigipilv) 2.0 is focused on the microservice architecture and full automation of the deployment processes. This means that no human resources are needed to maintain the systems – the client bears full responsibility. The SDN allows for the integration of the client's network into a distributed state cloud (Riigipilv). The concept paper on Riigipilv 2.0 is currently being developed and will be published later in 2022. The implementation is scheduled for the first half of 2023.

AI use cases and solutions would utilise the state cloud (Riigipilv) 2.0 as the target technical infrastructure to manage the needs throughout the AI development lifecycle, such as launching the servers and resources to start experiments and deploying the finalised solution in a production environment. Figure 13 presents a target reference technology architecture with core AI components that can be integrated with other technical components. The following architecture includes important AI components also in terms of enabling MLOps practices.

Figure 13. Target reference technology architecture with core AI components with other components



5.1.2 Data and information

Data is the most critical element to consider in providing AI services. In chapter 2, we analysed and described the data sources and public services in detail.

This section extends the explanation and highlights how the integrated data layers and AI services would impact each other, as well as the main AI dependencies to the data models. The aspects of data models that have the greatest impact on AI solutions are as follows:

- Data storage
- Data quality
- Data volume
- Data variation (less imbalance data distribution)
- Data availability
- Data accessibility
- Data attributes (data features to be used in AI development)

All the above-mentioned aspects of data models affected the development and implementation of AI solutions. For example, if the data was of poor quality or the volume of data was small, or data availability and accessibility was low or there was a bias in data distributions, this would affect the performance of the AI model and AI service. Therefore, a robust data architecture with proper data management principles is necessary for the use and development of AI services.

During our process of evaluating AI use cases, the key factors related to aspects of the target architecture were considered. For example, the assessment considered different aspects such as data availability, data quality, data volume and data readiness to be utilised for AI modelling.

5.1.3 Organisation, culture & talent

This section explains how AI infrastructure impacts the existing DGE infrastructure in terms of the organisation's operating model. It was highlighted in the AS-IS analysis that several challenges were in place to drive AI initiatives and projects.

They were mainly related to the complexity of coordination and communication, lack of resources and in-house skills and capabilities, lack of streamlined practices and distribution of roles. Initiating advanced AI solutions without addressing the problems associated with the existing operating model would not only increase the hassle but could also result in project failure. Therefore, there was a need to design new roles and governance practices to lead AI use cases and solutions.

AIOps was one of the ways to understand how organisations could evolve their siloed IT operations to support advanced AI-enabled solutions. One of the major reasons was that many organisations failed to realise the difference between AI development and solution development.

For example, the solution development lifecycle started with the design of services and users, while AI development began with formulating and experimenting with AI models based on available data. AIOPs practices streamlined AI development activities within solution development and reduced the risk of failure by validating how AI can solve the problem first, and how AI components would be utilised throughout the solution.

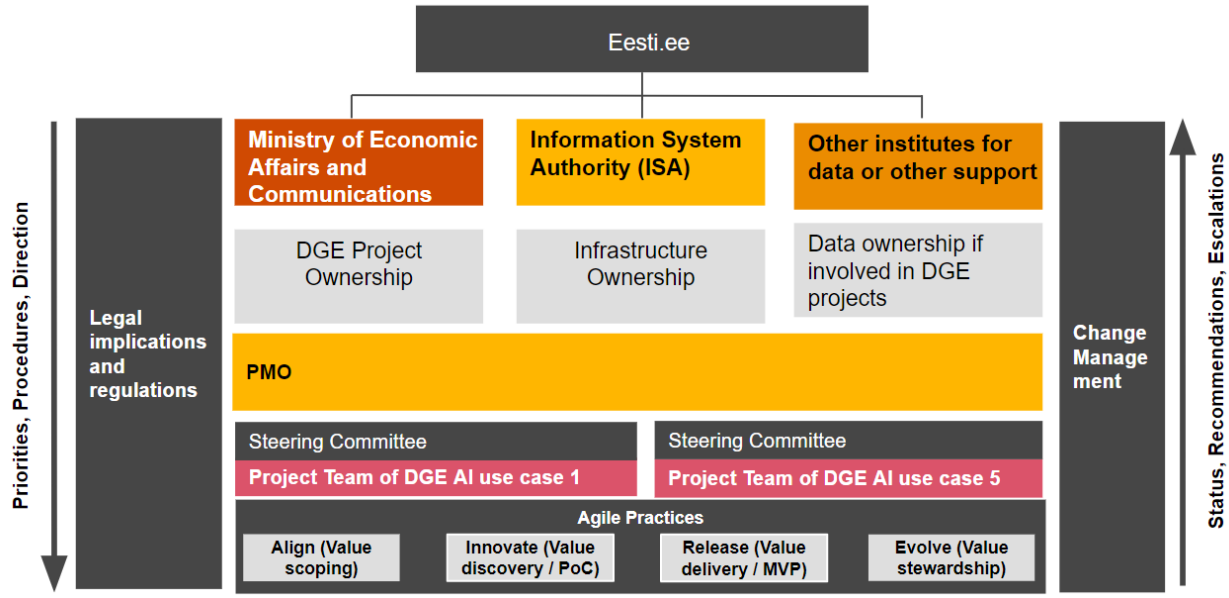
Moreover, the AI lifecycle impacted all levels of development: a) technical infrastructure b) data layer c) service layer (user-centric design). Therefore, implementing the right governance model at all these layers for AI projects would result in successful AIOPs practices.

One way was to use RACI rules to clearly define which organisation and team would be responsible and accountable for which activities. For example, if the infrastructure support was to be provided by ISA, it would be important to discuss and plan resources across teams and organisations. Some of the resources that needed to be discussed were the following:

- AI sandbox environments.
- Proper versioning mechanism for AI models and datasets.
- MLOPs practices for deployments and rollouts.

Furthermore, Figure 14 presents a target governance structure that covers important organisational aspects. For example, to ensure the development of the selected AI use case, change management and legal aspects need to be addressed from bottom to top and vice versa. Various forms of PMO ownership and practices are already in place. However, using agile practices tailored to the stages of the AI development lifecycle could improve the existing structure.

Figure 14. Target governance structure of the Digital state portal which incorporates AI solutions



5.1.4 Process and Integration

Digital gateway processes for enterprises would become more complex when coordination and communication with other organisations and public institutes were required. The right level of segregation in AI operating practices would drive coherent integration of processes and functions. It was important to clearly design the processes required to apply the AI operating model (AIOps). These processes and their underlying functions would initiate and drive activities in the right direction using the right capabilities.

The main activities would involve data sharing across different public institutes, storing data in a machine-readable format, and then utilising the data to develop and integrate AI models. The implementation of these activities impacted different layers and process levels which had to be considered:

Integrating AI required the enhancement of technical infrastructure, especially AI-related infrastructure such as GPUs. Given the abundance of tools and the lack of real best practices, AI infrastructure could easily become chaotic, hindering the development and scaling of AI services and solutions.

Data acquisition processes required a forward-looking approach that looked beyond data integration and one-time data transmission to consider ways to improve data flows to capture more data for AI models. For example, adding a feature to capture entrepreneurs' favourite topics provided important information when developing AI-based recommendations.

Furthermore, ethical AI development and processes related to trustworthiness needed to be defined according to the specific AI solution requirements. The misuse of AI could lead to significant reputational damage and financial costs for governments. Such risks could be minimized by using ethical approaches to the development of machine learning models.

Utilising explainable machine learning models reduced risks and was the best option to use where possible. There are different AI explainable methods, e.g., SHAP[30], that could also be used to explain the results of AI prediction. Using explainable methods in the development process also enhanced ethical and responsible artificial intelligence practices.

[30] SHAP: https://shap.readthedocs.io/en/latest/index.html

5.1.5 User-centric design

The user-centric design and integration at the service level required first and foremost user experience design and planning, where user profiling can be developed using AI. For example, developing user interfaces where entrepreneurs could store their favourite information and see personalised recommendations.

When designing, developing and integrating AI functionalities on a government portal, it was necessary to take into account the principle of service and user interface design.

The AIOps processes have impacted service design elements specifically, as defining user journeys where AI predictions could be used changes the existing user journeys. For example, the current user journey had the following main steps:

1. Entrepreneurs log in to the portal.
2. Entrepreneurs can view their company details.
3. Entrepreneurs can discover different information, such as financing information.
4. Entrepreneurs would be redirected to specific sites to apply, for example, for public sector financing.
5. Entrepreneurs can submit forms and see the status on the state portal

Selected use cases and integrating AI solutions into the services of the state portal would enhance the current user experience and journeys. For example, the selected DGE AI use case “personalised recommendations for entrepreneurs” would require at least the following additional steps integrated with the current user journeys:

- Entrepreneurs need to be asked what information they are most interested in.
- Entrepreneurs can mark information as favourites.

5.1.6 Legal

Using AI solutions in the DGE environment also impacted the legal frameworks associated with it. During the AS-IS analysis, it was identified that there were legislative and procedural shortcomings among public institutions, especially when it came to data sharing, data coordination and AI projects.

The readiness and willingness of institutions were among the factors that contribute to it. In general, the Estonian AI Strategy 2020-2023, Data Protection and AI regulation at the European Union level basically defined all legislative procedures and regulatory requirements.

Due to the developed e-society in Estonia, a multitude of laws have permitted the use of automated data processing in various tasks assigned to the public sector. Although there was still a lack of relevant case law, data and AI-related laws and regulations should be comprehensively analysed when implementing an AI use case.

Below is an overview of relevant Estonian laws and regulations[31] for the implementation of AI systems. These legal aspects were applicable to the implementation of any of the selected use cases[32]:

Public Information Act (PIA): The purpose of the PIA is to ensure that every person can access information intended for public use. Public authorities must provide the public with the opportunity to monitor the performance of public duties. Public authorities must ensure clarity and disclosure of information concerning the provision of public services. Public authorities must also disclose information on the purpose, scope and methods of personal data processing, transfer of personal data to third parties, including other agencies, and provision of access to personal data to the public, as well as on the right and procedure for familiarising a person with data concerning them personally.

[31] Estonia state liability acts https://www.riigiteataja.ee/akt/783043?leiaKehtiv

[32] Estonia Country research on Fundamental Rights https://fra.europa.eu/sites/default/files/fra_uploads/fra-ai-project-estonia-country-research_en.pdf

Administrative Procedure Act: The purpose of the act is to ensure a uniform administrative procedure that allows for the participation of interested parties and judicial control. The act is relevant when issuing administrative acts using AI technologies. Although AI is not specifically mentioned. This act states that fundamental rights and freedoms or other subjective human rights may be restricted only pursuant to the law; administrative acts and measures shall be appropriate, necessary and proportionate to the stated objectives.

Personal Data Protection Act (PDPA): The PDPA protects individuals in the processing of personal data and provides standards for implementing the GDPR and transposing Directive (EU) 2016/680. It also stipulates the method of state supervision over the processing of personal data and the liability for violations of requirements thereof.

Protection of personal data: The act sets the requirements for the maintenance of networks and information systems necessary for the functioning of society. These requirements also apply to state and local government networks and information systems. It also establishes legal obligations to prevent and resolve cyber incidents. AI systems developed and implemented by the state must comply with the cybersecurity requirements established in this act to ensure the protection of the personal data of all Estonian citizens.

Law of Obligations Act: This act establishes liability for damage caused by a major source of danger, regardless of the person's culpability. A matter or activity is designated as a major source of danger if it can cause major or frequent damage, even if it is handled or dealt with by a specialist.

6. Implementation roadmap for selected AI use cases

The core elements of the AI implementation roadmap consisted of several key considerations that aligned technical implementation with the organisation's objectives and solution ownership. These core elements aligned with AI design, development, validations, deployment, monitoring, and maintenance, supporting the organisation's vision.

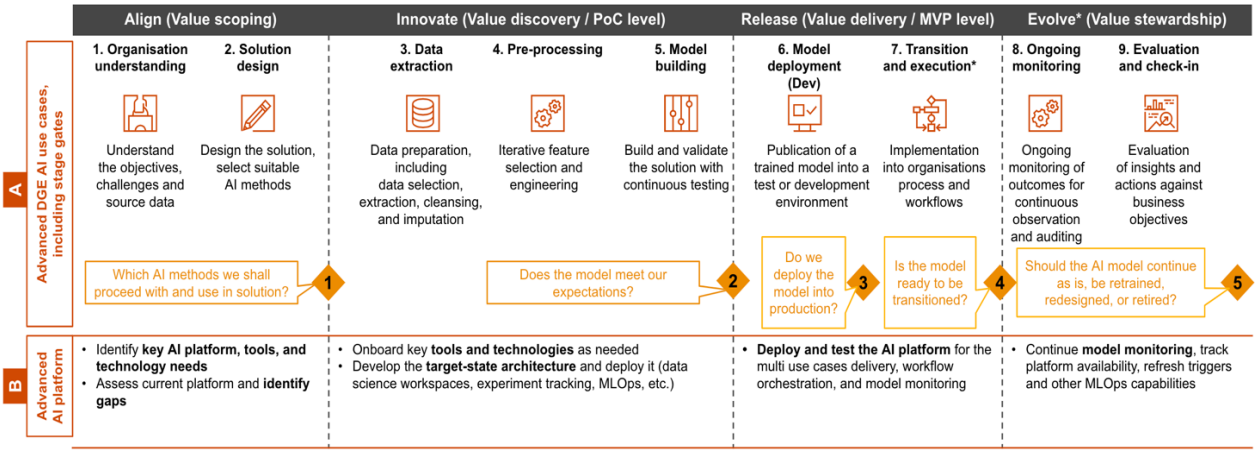
Moreover, agile practices that facilitated end-2-end implementation enhanced the overall process and added flexibility, adaption, and growth. Agile and Integrated ways of working required a scalable framework that incorporated the implementation aspects that were important for the growth and sustainability of AI solutions.

PwC BxT Agile framework is a generalised agile-based framework that offers a multi-phase agile approach to conceptualising, designing and defining implementation roadmaps. The main phases of the PwC BxT agile framework are:

1. **Align**: In this phase, the value scoping for AI use cases is assessed.
2. **Innovate:** In this phase, the focus is on understanding and discovering the value of AI use cases.
3. **Release:** In this phase, the focus is on creating value from the AI use cases.
4. **Evolve**: This phase aims to establish value management of the AI use cases.

Using the agile BxT framework, the following Figure 15 describes the end-2-end process of implementing DGE AI. The first step was to align the organisation's objectives and scoping value. Part of this process has already been assessed and evaluated in this report, e.g., the selection of AI use cases and knowing the organisation's objectives.

Figure 15. An overview of DGE AI use cases' development process



The second stage was the start of innovation and value discovery. This stage consisted of proof of concepts and validation of the applicability of AI models in the overall solution. The third stage was the solution release and value delivery. It consisted of developing and releasing a minimum viable product (MVP). The last stage was development, where value management was required. It involved further development and maintenance of the AI solution.

6.1 Implementation roadmaps for DGE selected use cases

Implementation roadmaps were prepared for both selected use cases which consisted of an implementation plan with key activities, required roles and a responsibility matrix using RACI, as well as an effort estimate for each activity to estimate development and maintenance costs.

The implementation roadmaps below used the DGE AI development process with the stages presented in Figure 15 above. As described above, these stages showed the phases of 1) scoping the solution, 2) proof of concept, 3) releasing a minimum viable product (MVP), and 4) developing and maintaining the solution. Implementation roadmaps used these four stages to define the essential development activities for each phase and then to map role requirements and estimations to the core activities.

Furthermore, the implementation roadmaps provided the activities to start the implementation. These activities especially related to data extraction and preparation were critical because AI solution depends on the quality and readiness of the data. As described in the data assessment section (chapter 2), most of the current data sources were identified as a low quality and with less AI readiness.

However, the listed data sources in the implementation plan were identified good enough to begin AI proof-of-concept phase. The main reason were that most of the listed data sources contained text information which can be validated by training language-based machine learning models, e.g. content-based, collaborative filtering.

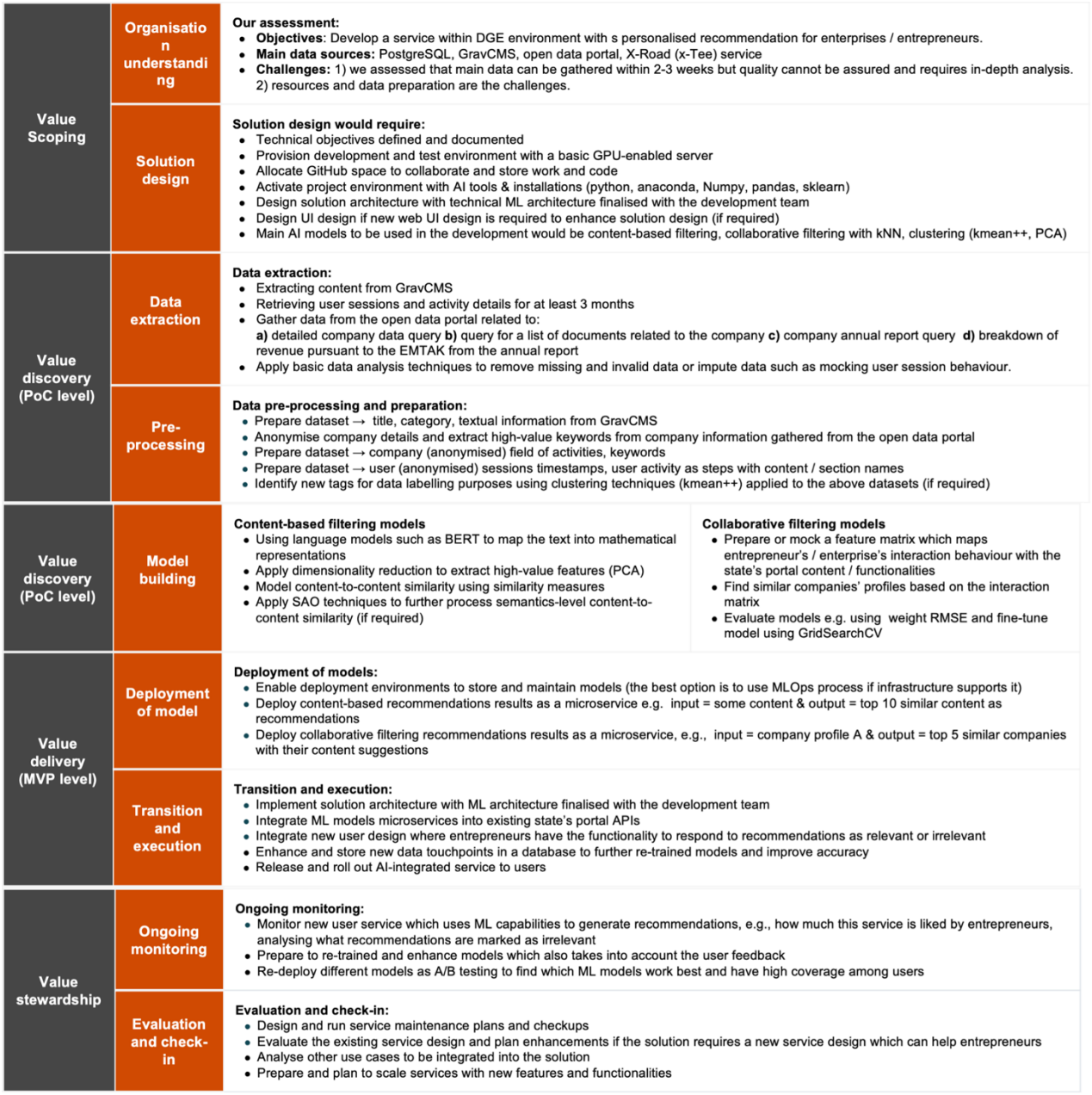
The machine learning model depends on the training datasets which can be created using highlighted data sources. The challenge-related to raw text can be overcome by applying the unsupervised learning techniques, e.g. kmeans++ or SAO (subject-action-object) NLP technique which is used to convert unstructured textual data into structured textual data. These unsupervised learning techniques can be used to categories data into clusters, and then custom labels can be assigned to each clusters which can be used in the training datasets.

Moreover, in case the unsupervised learning techniques fails to produce ideal results, then even poor results can be used to investigate the data elements which can be enriched either manually or automatically if possible. The enriched data can be used to produce quality training dataset to train efficient machine learning models.

6.1.1 Use case "personalised recommendations"

The use case aimed to use information from the DGE portal related to enterprises, user activity, e.g., what entrepreneurs are currently searching for on the portal, and other sources to generate personalised recommendations.

Figure 16. Implementation plan with activities for DGE's selected use case #1 "personalised recommendations"



This use case aimed to provide recommendations that would be dynamically adjusted based on recent user activity and views on the portal. The result would be an improvement of the Digital Gateway for Enterprises (DGE) service by providing personalised recommendations based on the entrepreneur's latest activities on the portal.

Responsibility matrix for the use case "personalised recommendation"

Figure shows the matrix of role requirements and responsibilities for the use case "personalised recommendation" in the DGE portal.

Figure 17. Required roles and responsibilities matrix for DGE's selected use case #1 "personalised recommendations"

Stages	Steps	Activities	PM	Data Scientists	Data/ML Engineers	Technical Lead	UI Designer	Web Developers
Value Scoping	Solution Design	Technical objectives defined and documented	R	C	C	A	C	C
		Provision development and test environment with a basic GPU-enabled server	I			RA		
		Allocate GitHub space to collaborate and store work and code	I		A	RA		
		Activate project environment with AI tools & installations (python, anaconda, Numpy, pandas, sklearn)	I	C	C	RA		
		Design solution architecture with technical ML architecture finalised with the development team		C	C	RA	C	C
		Design UI design if new web UI design is required to enhance solution design (if required)	R			C	A	C
Value Discovery (PoC level)	Data Extraction	Extracting content from GravCMS			RA	CI		
		Retrieving user sessions and activity details for at least 3 months			RA	CI		
		Gather data from the open data portal related to: **a)** detailed company data query **b)** query for a list of documents related to the company **c)** company annual report query **d)** breakdown of revenue pursuant to the EMTAK from the annual report			RA	CI		
		Apply basic data analysis techniques to remove missing and invalid data or impute data such as mocking user session behaviour.		RA		C		
	Data Pre-processing	Prepare dataset → title, category, textual information from GravCMS		RA		I		
		Anonymise company details and extract high-value keywords from company information gathered from the open data portal		RA		I		
		Prepare dataset → company (anonymised) field of activities, keywords		RA		I		
		Prepare dataset → user (anonymised) sessions timestamps, user activity as steps with content / section names		RA		I		
		Identify new tags for data labelling purposes using clustering techniques (kmean++) applied to the above datasets (if required)		RA	C	I		
Value Discovery (PoC level)	Model Building	Plan to build and maintain versions of content-based and collaborative filtering recommendations models	I	A	C	R		
		Using language models such as BERT to map textual into mathematical representations		RA				
		Apply dimensionality reduction to extract high-value features (PCA)		RA				
		Model content-to-content similarity using similarity measures		RA		C		
		Apply SAO techniques to further process semantic-level content-to-content similarity (if required)		RA		C		
		Prepare or mock a feature matrix which maps entrepreneur's / enterprise's interaction behaviour with the state's portal content / functionalities		RA		C		
		Find similar companies' profiles based on the interaction matrix		RA		C		
		Evaluate models, e.g. using weight RMSE and fine-tune model using GridSearchCV		RA		CI		
		Document code and results and present them to the team.	I	RA		C		
Value Delivery (MVP level)	Deployment of Models	Enable deployment environments to store and maintain models (the best option is to use MLOps process if infrastructure supports it)	I		RA	C		
		Deploy content-based recommendations results as a microservice, e.g. input = some content & output = top 10 similar content as recommendations			RA	C		
		Deploy collaborative filtering recommendations results as a microservice, e.g. input = company profile A & output = top 5 similar companies with their content suggestions			RA	C		
Value Delivery (MVP level)	Transition and execution	Implement solution architecture with ML architecture finalised with the development team				R	A	A
		Integrate ML models microservices into existing state's portal APIs			C	R		A
		Enhance and store new data touchpoints in a database to further re-trained models and improve accuracy				R		A
		Release and roll out AI-integrated service to users	CI			R		A
Value Stewardship	Ongoing monitoring	Monitor new user service which uses ML capabilities to generate recommendations, e.g. how much this service is liked by entrepreneurs, analysing what recommendations are marked as irrelevant	R	C		A	C	
		Prepare to re-trained and enhance models which also takes into account the user feedback	I	R	C	A		
		Re-deploy different models as A/B testing to find which ML models work best and have high coverage among users	I		R	RA		R
	Evaluation and check in	Design and run service maintenance plans and checkups	I	C	C	RA		C
		Evaluate the existing service design and plan enhancements if the solution requires a new service design which can help entrepreneurs	C			RA	C	
		Analyse other use cases to be integrated into the solution	RA	C		C	C	
		Prepare and plan to scale services with new features and functionalities	I	C	C	RA	C	C

The responsibility matrix or RACI model in Figure shows the activities in the rows and the required and involved roles in the columns. The responsibility levels were divided into the following categories:

R – The performer (responsible party) performs an activity that could have multiple tasks. For each activity, at least one person was assigned to be responsible for that activity. Responsibility could also be shared. The number of persons fulfilling the responsibility was determined by the responsible person.

A – The accountable *party* was responsible for the implementation of the action plan. It indicated that a person had to manage and control the work of the executors. There should not have been more than one person responsible for a single activity.

C – The consulted party was a *party* that shared information and/or gave advice to responsible persons and executors. Consultants were usually experts or individuals who could be directly affected by the work results and were communicated with throughout the work process.

I - The informed party was a party that did not have a specific activity or task in the work process. However, they were informed about decisions and work results. Information exchange generally took place after the activity was completed. Informants were executors or responsible persons.

Effort estimation for the use case "personalised recommendation"

Figure presents effort estimation for the use case "personalised recommendation" in the DGE portal. Each activity listed in the rows has been assessed for its expected scope and priority. The assessment of the scope was based on the full-time equivalent on the following scale:

XS – it took a few days to complete the work; we assumed that it would be roughly 2 days.
S – it took a few weeks to complete the work; we assumed that it would be roughly 11 days.
M – it took a few months to complete the work; we assumed that it would be roughly 27 days.

Figure 18. Estimated effort for DGE's selected use case #1 "personalised recommendations"

Stages	Steps	Activities	1.0 FTE PM	2.0 FTE Data Scientists	1.0 FTE Data/ML Engineers	1.0 FTE Technical Lead	1.0 FTE UI Designer	2.0 FTE Web Developers	Effort Estimation	Total FTE
Value Scoping	Solution Design	Technical objectives defined and documented	R	C	C	A	C	C	XS	4.0
		Provision development and test environment with a basic GPU-enabled server	I			RA			XS	2.0
		Allocate GitHub space to collaborate and store work and code	I		A	RA			XS	3.0
		Activate project environment with AI tools & installations (python, anaconda, Numpy, pandas, sklearn)	I	C	C	RA			XS	2.0
		Design solution architecture with technical ML architecture finalised with the development team		C	C	RA	C	C	S	2.0
		Design UI design if new web UI design is required to enhance solution design (if required)	R			C	A	C	S	2.0
Value Discovery (PoC level)	Data Extraction	Extracting content from GravCMS			RA	CI			S	2.0
		Retrieving user sessions and activity details for at least 3 months			RA	CI			S	2.0
		Gather data from the open data portal related to: **a)** detailed company data query **b)** query for a list of documents related to the company **c)** company annual report query **d)** breakdown of revenue pursuant to the EMTAK from the annual report			RA	CI			S	2.0
		Apply basic data analysis techniques to remove missing and invalid data or impute data such as mocking user session behaviour.		RA		C			S	3.0
	Data Pre-processing	Prepare dataset → title, category, textual information from GravCMS		RA		I			S	3.0
		Anonymise company details and extract high-value keywords from company information gathered from the open data portal		RA		I			XS	3.0
		Prepare dataset → company (anonymised) field of activities, keywords		RA		I			XS	3.0
		Prepare dataset → user (anonymised) sessions timestamps, user activity as steps with content / section names		RA		I			XS	3.0
		Identify new tags for data labelling purposes using clustering techniques (kmean++) applied to the above datasets (if required)		RA	C	I			S	3.0
Value Discovery (PoC level)	Model Building	Plan to build and maintain versions of content-based and collaborative filtering recommendations models	I	A	C	R			S	4.0
		Using language models such as BERT to map textual into mathematical representations		RA					S	2.0
		Apply dimensionality reduction to extract high-value features (PCA)		RA					S	2.0
		Model content-to-content similarity using similarity measures		RA		C			S	3.0
		Apply SAO techniques to further process semantic-level content-to-content similarity (if required)		RA		C			S	3.0
		Prepare or mock a feature matrix which maps entrepreneur's / enterprise's interaction behaviour with the state's portal content / functionalities		RA		C			S	3.0
		Find similar companies' profiles based on the interaction matrix		RA		C			S	3.0
		Evaluate models, e.g. using weight RMSE and fine-tune model using GridSearchCV		RA		CI			S	3.0
		Document code and results and present them to the team.	I	RA		C			S	4.0
Value Delivery (MVP level)	Deployment of Models	Enable deployment environments to store and maintain models (the best option is to use MLOps process if infrastructure supports it)	I		RA	C			M	3.0
		Deploy content-based recommendations results as a microservice, e.g. input = some content & output = top 10 similar content as recommendations			RA	C			S	2.0
		Deploy collaborative filtering recommendations results as a microservice, e.g. input = company profile A & output = top 5 similar companies with their content suggestions			RA	C			S	2.0
Value Delivery (MVP level)	Transition and execution	Implement solution architecture with ML architecture finalised with the development team				R	A	A	M	3.0
		Integrate ML models microservices into existing state's portal APIs			C	R		A	M	3.0
		Enhance and store new data touchpoints in a database to further re-trained models and improve accuracy				R		A	M	2.0
		Release and roll out AI-integrated service to users	CI			R		A	M	2.0
Value Stewardship	Ongoing monitoring	Monitor new user service which uses ML capabilities to generate recommendations, e.g. how much this service is liked by entrepreneurs, analysing what recommendations are marked as irrelevant	R	C		A	C		S	4.0
		Prepare to re-trained and enhance models which also takes into account the user feedback	I	R	C	A			M	3.0
		Re-deploy different models as A/B testing to find which ML models work best and have high coverage among users	I		R	RA		R	M	4.0
	Evaluation and check in	Design and run service maintenance plans and checkups	I	C	C	RA		C	M	3.0
		Evaluate the existing service design and plan enhancements if the solution requires a new service design which can help entrepreneurs	C			RA	C		M	2.0
		Analyse other use cases to be integrated into the solution	RA	C		C	C		M	2.0
		Prepare and plan to scale services with new features and functionalities	I	C	C	RA	C	C	M	4.0

Cost estimation for the use case " personalised recommendation"

The following cost estimations were based on the implementation plan activities, responsibility matrix, effort estimations and the total number of full-time equivalents (FTE). The following assumptions were made to provide an estimate of the cost of the MVP solution. The maintenance cost is assumed to be 10 percent of the implementation cost because it considered only maintenance aspects such as infrastructure costs, monitoring, and minor fixes.

- A full-time employee works 40 hours a week and 8 hours a day (daily work hours)
- The average cost per development hour is approximately €85/per hour.
- A formula to derive implementation costs: Days * Daily work hours * Sum of total FTEs for all activities * Hourly rate.
- A formula to derive maintenance costs: implementation cost * 0.10. The maintenance cost is assumed to be 10 percent of the implementation cost because it considered only maintenance aspects such as infrastructure costs, monitoring, and minor fixes.

Table 6. Cost estimation for DGE's selected use case #1 "personalised recommendations"

Phase	Implementation cost	Maintenance cost
MVP	€639 880	€63 988

6.1.2 Use case "smart suggestions for deadlines and obligations"

The use case aimed to use AI to enhance an existing solution. It aimed to process multiple data sources and suggest deadlines and obligations to entrepreneurs that arises from the nature of their field of activity and other relevant factors.

Figure presents an AI implementation plan with key activities for the use case "smart suggestions for deadlines and suggestions" on the DGE portal.

Figure 19. Implementation plan with activities for DGE selected use case #5 "smart suggestions for deadlines and obligations"

Value Scoping	Organisation understanding	**Our assessment:** • **Objectives**: Develop a service within the DGE environment which provides smart deadlines and obligations to enterprises / entrepreneurs. • **Main data sources:** Open data portal, Riigi Teataja, economic register (MTR), registers and information systems (Registrite ja Infosüsteemide) • **Challenges:** 1) we assessed that partially the data can be gathered in a few weeks but quality cannot be assured and requires in-depth analysis. 2) resources and data preparation are the challenges.
	Solution design	**Solution design would require:** • Technical objectives defined and documented • Provision development and test environment with basic a GPU-enabled server • Allocate GitHub space to collaborate and store work and code • Activate project environment with AI tools & installations (python, anaconda, Numpy, pandas, sklearn) • Design solution architecture with technical ML architecture finalised with the development team • Design UI design if new web UI design is required to enhance solution design (if required) • Main AI models to be used in the development would be content-based filtering, collaborative filtering with kNN, clustering (kmean++, PCA)
Value discovery (PoC level)	Data extraction	**Data extraction:** • Gather enterprise information from open data portal • Extract companies' details particularly related to field of activities that companies are operating in • Extract companies' details particularly related to previous deadlines and a list of known obligations (if applicable) • Gather documents from Registers and Information systems (Registrite ja Infosüsteemide) related to obligations as per field of businesses and other factors • Apply basic data analysis techniques to remove missing and invalid data
	Pre-processing	**Data pre-processing and preparation:** • Prepare dataset 01 →, e.g. company info (anonymised), field of activities, age, list of known obligations, no. of completed deadlines, list of applied licenses • Extract top keywords from pdf or other obligations documents for each field of activities • Prepare dataset 02 →, e.g. field of activities, top keywords extracted from obligations documents (if required) • Apply clustering techniques to categorise on the high level the similar field of activities with respect to obligations (top keywords extracted from obligations documents (if required)
Value discovery (PoC level)	Model building	**Collaborative filtering models** • Plan to build and maintain versions of collaborative filtering models • Prepare several feature matrices from dataset 01 to apply multiple-level collaborative filtering algorithms, e.g. finding the top 3 similarity vectors for each column from from dataset 01 • Rank feature similarity vectors to categories of similar companies • Find similar company profiles based on the deadlines and obligations • Evaluate models e.g., using weight RMSE and fine-tune model using GridSearchCV • Document code and results and present them to the team
Value delivery (MVP level)	Deployment of model	**Deployment of models:** • Enable deployment environments to store and maintain models (the best option is to use MLOps process if infrastructure supports it) • Deploy collaborative filtering model results as a microservice, e.g. input = company profile A & output = top 5 similar companies with their deadlines and obligations suggestions
	Transition and execution	**Transition and execution:** • Implement solution architecture with ML architecture finalised with the development team • Integrate ML models microservices into existing state's portal APIs • Integrate new user design where entrepreneurs have the functionality to respond to suggestions as relevant or irrelevant • Enhance and store new data touchpoints in a database to further re-trained models and improve accuracy • Release and roll out AI-integrated service to users
Value stewardship	Ongoing monitoring	**Ongoing monitoring:** • Monitor new user service which uses ML capabilities to generate suggestions, e.g. how much this service is liked by entrepreneurs, analysing what suggestions are marked as irrelevant • Prepare to re-trained and enhance models which also takes into account the user feedback • Re-deploy different models as A/B testing to find which ML models work best and have high coverage among users
	Evaluation and check-in	**Evaluation and check-in:** • Design and run service maintenance plans and checkups • Evaluate the existing service design and plan enhancements if the solution requires a new service design which can help entrepreneurs • Analyse other use cases to be integrated into the solution • Prepare and plan to scale services with new features and functionalities

The outcome of this use case would enhance the Digital Gateway for Enterprises (DGE) service, which would provide an overview of deadlines and obligations. The new service would use AI to notify entrepreneurs about different suggestions related to deadlines and obligations relevant to factors such as their line of business.

Responsibility matrix for the use case "smart suggestions for deadlines and obligations"

Figure presents the matrix of role requirements and responsibilities for the use case "smart suggestions for deadlines and obligations" in the DGE portal.

Figure 20. Required roles and responsibilities matrix for DGE's selected use case #5 "smart suggestions for deadlines and obligations"

Stages	Steps	Activities	PM	Data Scientists	Data/ML Engineers	Technical Lead	UI Designer	Web Developers
Value Scoping	Solution Design	Technical objectives defined and documented	R	C	C	A	C	C
		Provision development and test environment with a basic GPU-enabled server	I			RA		
		Allocate github space to collaborate and store work and code	I		A	RA		
		Activate project environment with AI tools & installations (python, anaconda, Numpy, pandas, sklearn)	I	C	C	RA		
		Design solution architecture with technical ML architecture finalised with the development team		C	C	RA	C	C
		Design UI design if new web UI design is required to enhance solution design (if required)	R			C	A	C
Value Discovery (PoC level)	Data Extraction	Gather enterprises' information from an open data portal			RA	CI		
		Extract companies' details particularly related to the field of activities that companies are operating in			RA	CI		
		Extract companies' details particularly related to previous deadlines and a list of known obligations (if applicable)			RA	CI		
		Gather documents from Registers and Information systems (Registrite ja Infosüsteemide) related to obligations as per field of businesses and other factors		RA		C		
		Apply basic data analysis techniques to remove missing and invalid data.		RA				
	Data Pre-processing	Prepare dataset 01 →, e.g. company info (anonymised), field of activities, age, list of known obligations, no. of completed deadlines, list of applied licenses		RA		I		
		Extract top keywords from pdf or other obligations documents for each field of activities		RA		I		
		Prepare dataset 02 →, e.g. field of activities, top keywords extracted from obligations documents (if required)		RA		I		
		Apply clustering techniques to categorise on the high level the similar field of activities with respect to obligations (top keywords extracted from obligations documents (if required)		RA		I		
Value Discovery (PoC level)	Model Building	Plan to build and maintain versions of collaborative filtering models	I	A	C	R		
		Prepare several feature matrices from dataset 01 to apply multiple-level collaborative filtering algorithms , e.g. finding the top 3 similarity vectors for each column from from dataset 01		RA				
		Rank feature similarity vectors to categories of similar companies		RA				
		Find similar company profiles based on the deadlines and obligations		RA		C		
		Evaluate models, e.g. using weight RMSE and fine-tune model using GridSearchCV		RA		C		
		Document code and results and present them to the team	I	RA		C		
Value Delivery (MVP level)	Deployment of Models	Enable deployment environments to store and maintain models (the best option is to use MLOps process if infrastructure supports it)	I		RA	C		
		Deploy collaborative filtering model results as a microservice, e.g. input = company profile A & output = top 5 similar companies with their deadlines and obligations suggestions			RA	C		
	Transition and execution	Implement solution architecture with ML architecture finalised with the development team				R	A	A
		Integrate ML models microservices into existing state's portal APIs			C	R		A
		Integrate new user design where entrepreneurs have the functionality to respond to suggestions as relevant or irrelevant	I			RA	A	A
		Enhance and store new data touchpoints in a database to further re-trained models and improve accuracy				R		A
		Release and roll out AI-integrated service to users	CI			R		A
Value Stewardship	Ongoing monitoring	Monitor new user service which uses ML capabilities to generate suggestions, e.g. how much this service is liked by entrepreneurs, analysing what suggestions are marked as irrelevant	R	C		A	C	
		Prepare to re-trained and enhance models which also takes into account the user feedback	I	R	C	A		
		Re-deploy different models as A/B testing to find which ML models work best and have high coverage among users	I		R	RA		R
Value Stewardship	Evaluation and check in	Design and run service maintenance plans and checkups	I	C	C	RA		C
		Evaluate the existing service design and plan enhancements if the solution requires a new service design which can help entrepreneurs	C			RA	C	
		Analyse other use cases to be integrated into the solution	RA	C		C	C	
		Prepare and plan to scale services with new features and functionalities	I	C	C	RA	C	C

Estimated effort for the use case "smart suggestions for deadlines and obligations"

Figure presents the effort estimate for the use case "smart suggestions for deadlines and obligations" in the DGE portal. Each activity listed in the rows has been assessed for its expected scope and priority. The assessment of the scope was based on a full-time equivalent on the following scale:

XS – it took a few days to complete the work; we assumed that it would be roughly 2 days
S – it took a few weeks to complete the work; we assumed that it would be roughly 11 days
M – it took a few months to complete the work; we assumed that it would be roughly 27 days

Figure 21. Estimated effort for DGE's selected use case #5 "smart suggestions for deadlines and obligations"

Stages	Steps	Activities	1.0 FTE PM	2.0 FTE Data Scientists	1.0 FTE Data/ML Engineers	1.0 FTE Technical Lead	1.0 FTE UI Designer	2.0 FTE Web Developers	Effort Estimation	Total FTE
Value Scoping	Solution Design	Technical objectives defined and documented	R	C	C	A	C	C	XS	4.0
		Provision development and test environment with a basic GPU-enabled server	I			RA			XS	2.0
		Allocate github space to collaborate and store work and code	I		A	RA			XS	3.0
		Activate project environment with AI tools & installations (python, anaconda, Numpy, pandas, sklearn)	I	C	C	RA			XS	2.0
		Design solution architecture with technical ML architecture finalised with the development team		C	C	RA	C	C	S	2.0
		Design UI design if new web UI design is required to enhance solution design (if required)	R			C	A	C	S	2.0
Value Discovery (PoC level)	Data Extraction	Gather enterprises' information from an open data portal			RA	CI			S	2.0
		Extract companies' details particularly related to the field of activities that companies are operating in			RA	CI			S	2.0
		Extract companies' details particularly related to previous deadlines and a list of known obligations (if applicable)			RA	CI			S	2.0
		Gather documents from Registers and Information systems (Registrite ja Infosüsteemide) related to obligations as per field of businesses and other factors		RA		C			S	3.0
		Apply basic data analysis techniques to remove missing and invalid data.		RA					S	3.0
	Data Pre-processing	Prepare dataset 01 →, e.g. company info (anonymised), field of activities, age, list of known obligations, no. of completed deadlines, list of applied licenses		RA		I			S	3.0
		Extract top keywords from pdf or other obligations documents for each field of activities		RA		I			XS	3.0
		Prepare dataset 02 →, e.g. field of activities, top keywords extracted from obligations documents (if required)		RA		I			XS	3.0
		Apply clustering techniques to categorise on the high level the similar field of activities with respect to obligations (top keywords extracted from obligations documents (if required)		RA		I			XS	3.0
Value Discovery (PoC level)	Model Building	Plan to build and maintain versions of collaborative filtering models	I	A	C	R			S	4.0
		Prepare several feature matrices from dataset 01 to apply multiple-level collaborative filtering algorithms , e.g. finding the top 3 similarity vectors for each column from from dataset 01		RA					S	2.0
		Rank feature similarity vectors to categories of similar companies		RA					S	2.0
		Find similar company profiles based on the deadlines and obligations		RA		C			S	2.0
		Evaluate models, e.g. using weight RMSE and fine-tune model using GridSearchCV		RA		C			S	3.0
		Document code and results and present them to the team	I	RA		C			S	4.0
Value Delivery (MVP level)	Deployment of Models	Enable deployment environments to store and maintain models (the best option is to use MLOps process if infrastructure supports it)	I		RA	C			M	3.0
		Deploy collaborative filtering model results as a microservice, e.g. input = company profile A & output = top 5 similar companies with their deadlines and obligations suggestions			RA	C			S	2.0
	Transition and execution	Implement solution architecture with ML architecture finalised with the development team				R	A	A	M	3.0
		Integrate ML models microservices into existing state's portal APIs			C	R		A	M	3.0
		Integrate new user design where entrepreneurs have the functionality to respond to suggestions as relevant or irrelevant	I			RA	A	A	S	2.0
		Enhance and store new data touchpoints in a database to further re-trained models and improve accuracy				R		A	M	2.0
		Release and roll out AI-integrated service to users	CI			R		A	M	2.0
Value Stewardship	Ongoing monitoring	Monitor new user service which uses ML capabilities to generate suggestions, e.g. how much this service is liked by entrepreneurs, analysing what suggestions are marked as irrelevant	R	C		A	C		S	4.0
		Prepare to re-trained and enhance models which also takes into account the user feedback	I	R	C	A			M	3.0
		Re-deploy different models as A/B testing to find which ML models work best and have high coverage among users	I		R	RA		R	M	4.0
Value Stewardship	Evaluation and check in	Design and run service maintenance plans and checkups	I	C	C	RA		C	M	3.0
		Evaluate the existing service design and plan enhancements if the solution requires a new service design which can help entrepreneurs	C			RA	C		M	2.0
		Analyse other use cases to be integrated into the solution	RA	C		C	C		M	2.0
		Prepare and plan to scale services with new features and functionalities	I	C	C	RA	C	C	M	4.0

Cost estimation for the use case "smart suggestions for deadlines and obligations"

The following cost estimations were derived based on the effort estimates presented in Figure . The following assumptions have been made to arrive at the estimated cost of the MVP solution:

- A full-time employee works 40 hours a week and 8 hours a day (daily work hours).
- The average cost per development hour is approximately €85/ per hour.

- A formula to derive implementation cost: Days * Daily work hours * Sum of total FTEs for all activities * Hourly rate.
- A formula to derive maintenance cost: implementation cost * 0.10. The maintenance cost is assumed to be 10 percent of the implementation cost because it considered only maintenance aspects such as infrastructure costs, monitoring, and minor fixes.

Table 7. Cost estimation for DGE's selected use case #1 "smart suggestions for deadlines and obligations"

Phase	Implementation cost	Maintenance cost
MVP	€542 640	€54 264

Sources

Materials used for the analysis

1) State portal eesti.ee, E-Residency: https://www.eesti.ee/en/doing-business/eresidency
2) E-Residency, Who should apply? https://www.e-resident.gov.ee/become-an-e-resident/
3) X-Road ISA, X-tee, What is X-tee? www.x-tee.ee
4) "Riigiportaali eesti.ee kasutaja rahulolu analüüs" (2020, RIA ja MKM)
5) DGE D3 State of play report
6) Open data e-business register https://avaandmed.ariregister.rik.ee/en
7) Digital Economy and Society Index (DESI) 2022, Estonia.
8) "Ettevõtte finantseerimise ärisündmuse analüüsi lõpparuanne"(EY)
9) "Riigiportaali-kasutatavuse-analuutika-loppraport-12.02.2021.pdf" (2021, Nortal)
10) “Ettevõtjate jaoks ühtse veebipõhise kontaktpunkti visioon “(2020, Nortal)
11) Krati plaan 2022-2023, https://en.kratid.ee/kratikava
12) RIA, https://www.ria.ee/en.html
13) RIA, https://www.ria.ee/en/state-information-system/state-portal-eestiee.html
14) Riigiportaali analüütika tutvustus_07.09.2022.pptx(shared by State's portal Data Analyst)
15) Tax and customs board https://www.emta.ee/en/business-client/e-services-training-courses/how-use-e-services/x-tee-services
16) e-Business register https://ariregister.rik.ee/eng
17) Statistics Estonia for developers https://www.stat.ee/en/statistikaamet/developers
18) Public data inquiries - Tax and Customs https://www.emta.ee/en/business-client/registration-business/inquiries-and-certificates/public-data-inquiries
19) Statistics Open data portal datasets related to Tax and Customs https://avaandmed.eesti.ee/information-holders/maksu-ja-tolliamet
20) Draft Act Information System (Eelnõude infosüsteem) https://www.evs.ee/en/general-principles
21) State Legislation Documents https://www.riigiteataja.ee/en/
22) Estonian Classification of Economic Activities (EMTAK) https://emtak.rik.ee/EMTAK/pages/klassifikaatorOtsing.jspx
23) Economic activities (*Majandustegevuse Register*) https://mtr.ttja.ee/?m=1
24) State Support Services Centre (*Riigi Tugiteenuste Keskus)* https://www.rtk.ee/asutus-uudised-kontakt/kontaktid/dokumendiregister
25) Agriculture Register and Information Board (Põllumajandus Registrite ja Informatsioini Amet) https://www.pria.ee/infokeskus
26) Centre for Environmental Investments (Keskkonnainvesteeringute Keskus) https://kik.ee/et/toetatavad-tegevused
27) Estonian Procurement Register (Riigihangete Resgitri) https://www.fin.ee/en/public-procurement-state-aid-and-assets/public-procurement-policy/public-procurement-register
28) Estonian Procurement Register (Riigihangete Resgitri) https://riigihanked.riik.ee/rhr-web/#/open-data
29) Tender electronic daily (TED) https://ted.europa.eu/TED/main/HomePage.do
30) Tax and custom board trainings https://www.emta.ee/en/business-client/e-services-training-courses/training-courses
31) Visit International training programs https://visiidid.ee/
32) Enterprise Estonia https://eas.ee/en/
33) European Union Agency for Fundamental Rights https://fra.europa.eu/en/project/2018/artificial-intelligence-big-data-and-fundamental-rights
34) Estonia Country research on Fundamental Rights https://fra.europa.eu/sites/default/files/fra_uploads/fra-ai-project-estonia-country-research_en.pdf
35) European Union AI watch https://ai-watch.ec.europa.eu/countries/estonia/estonia-ai-strategy-report_en
36) Explainable machine learning https://shap.readthedocs.io/en/latest/index.html

Appendices

6.2 Annex 1. Detail evaluation of the use cases

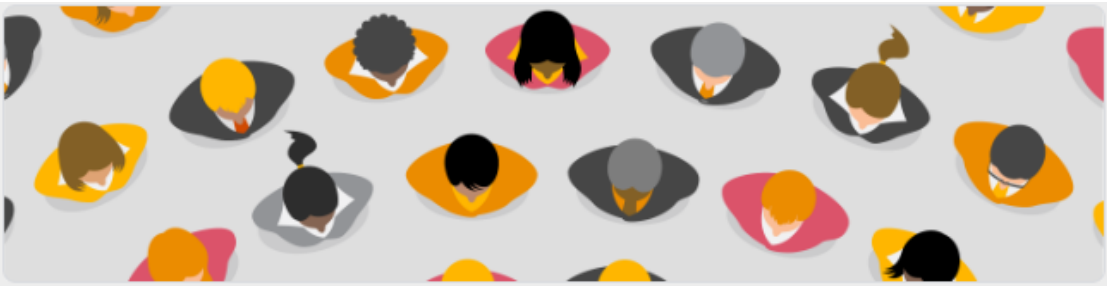
#	Use Case	Short Description	Perceived usefulness by entrepreneurs (n=31), scale 1-5	Perceived usefulness by Estonian Chamber of Commerce and Industry, scale 1-5	Expected Scalability	Expected benefit	no. of entrepreneurs affected	Expected frequency	Impact
1	Increase DGE users' engagement using personalized recommendations using AI	This use case intends to enhance personalization based on user recent activity on the portal. The aim is to develop an AI solution which uses session-based recommendations, which get dynamically adjusted based on user recent activities on the state portal.	3,42	3	2,5	3	3	3	2,875
11	Smart DGE search engine	This use case intends to tackle the challenge of applying smart AI based search on the DGE channel. An AI can be used to create the knowledge management system which possesses all levels of details related to Estonian fields of activities and business life-cycle events. Entrepreneurs can use this search to quickly search for any topic of interest.	4,2	4	2	2,5	3	3	2,625
3	Improve DGE process of providing tax and reporting information to entrepreneurs using AI	This use case intends to improve the process of providing most relevant content to the users especially related to taxes and reporting or other relevant factors based on the enterprises' field of activity and other related business events. The aim is to develop a solution which can predict and recommend information to entrepreneurs for the next steps, e.g., event driven report submissions.	4,19	4	2	2,5	3	2,5	2,5
5	Guide entrepreneurs through DGE smart suggestions channel related to deadlines and obligations	Enterprises work consistently to meet deadlines and obligations. This use case aims to use an AI solution to enhance the process of suggesting to entrepreneurs how they can move forward with their obligations.	4,65	4	2	2,5	3	2	2,375
4	Improve the DGE information channel for entrepreneurs to share and suggest business topic-based information using AI	This use case aims to use AI to solve the problem of finding relevant information from a wide range of topics related to business activities. The idea is to use advanced topic modelling techniques on unstructured data related to tax, obligations, compliance, and regulation etc. Entrepreneurs can subscribe to different topics and receive relevant information.	4,48	4	2	2	3	2	2,25
7	Improve suggestions related to public sector financing opportunities	This use case focuses on improving the process of sharing public sector financing opportunities to companies and entrepreneurs using an AI. The scope is to cover information related to co-financing, grants, loans, innovation funds, guarantees, and other related support offered by public services. Information concerning to procurements opportunities won't be cater in this use case. Furthermore, this use case would be able to take into accounts the information related to enterprises' financial status if available or shared by consent. Such information would enhance the AI-based suggestions for enterprises.	4,19	5	1,5	2	2	2	1,875

#	Use Case	Short Description	Perceived usefulness by entrepreneurs (n=31), scale 1-5	Perceived usefulness by Estonian Chamber of Commerce and Industry, scale 1-5	Expected Scalability	Expected benefit	no. of entrepreneurs affected	Expected frequency	Impact
2	Enhance DGE process of providing license and compliance information to entrepreneurs using AI	This use case intends to enhance the DGE channel to provide most relevant content to the users especially related to compliance and licenses based on the enterprises' field of activity. Activities depending on the nature of business contains different divisions, and each division consist of some compliance information to operate the business adhering to laws. Such information can be beneficial to business which requires compliance efforts to operate e.g., forestry, mining, manfucturing etc. The aim is to develop an AI-based solution which can take input related to nature of business and recommend information to entrepreneurs for the next steps, e.g., relevant activity and it's divisions which list down related compliance clauses and license required.	4,62	5	1,5	2,5	2	1	1,75
10	Improve the process of finding the right public sector experts	This use case aims to enhance through the DGE channel the process of finding the right public sector experts for difficult questions related to business activities. Entrepreneurs can share what they are looking for and an AI can be used to search in unstructured information and suggest the relevant experts and resources.	4,19	4	1	2	2	2	1,75
8	Smart search for finding EU level financing opportunities	This use case aims to improve the process of sharing EU level financing opportunities for Estonian entrepreneurs and enterprises. The focus is to extract information related to grants, funds, loans, and guarantees offered by EU and serve it to the Estonian companies to avail those financing opportunities. Processing information related to EU-level procurements would be out of scope for this use case.	4,29	2	1,5	2	1	2	1,625
6	Recommendations related to public sector trainings and events	This use case aims to enhance the process of sharing public sector related events (e.g., conferences, training) and information to entrepreneurs using AI. The AI with textual analysis can be done on the companies, entrepreneurs, and business activities data, and then suggesting relevant public sector events or articles.	3,63	2	1,5	1	2	2	1,625
9	Public procurement recommendations	This use case intends to help enterprises in the processes related to public procurement and tender submissions using an AI. A solution can use the previous submission history and give recommendations to enterprises to apply and participate in different public procurements.	4,15	4	1	1,5	1,5	2	1,5
12	Smart suggestions to improve business growth	This use case aims to provide support to enterprises who are exporting good and services. The focus is to help enterprises through smart analysis in order that they can enhance their current productivity and market share and expand to their services. The smart suggestions can be related to e.g., highlighting and suggesting that their export related field of activities had growth and in-general there are margins to capture by innovating and expanding business, and also on a high-level what kind of regulations and obligations would be necessary to move forward. The suggestions won't be related to finding grants, loans because there are other use cases and existing services which offer such support to entrepreneurs.	4,11	5	1,5	2	1,5	1,5	1,625

#	Use Case	Short Description	Data readiness	Technology		Infrastructure	MVP development cost		Ease of implementation
1	Increase DGE users' engagement using personalized recommendations using AI	This use case intends to enhance personalization based on user recent activity on the portal. The aim is to develop an AI solution which uses session-based recommendations, which get dynamically adjusted based on user recent activities on the state portal.	1,75	Data availability = medium =2 Data quality = low = 1 Data volume = high = 3 Data preparedness = low = 1	Algorithm simplicity = medium = 2	2	Data storage simplicity = medium = 2 Compute simplicity = medium = 2	Estimate hours >= 600 --> high	1,6875
11	Smart DGE search engine	This use case intends to tackle the challenge of applying smart AI based search on the DGE channel. An AI can be used to create the knowledge management system which possesses all levels of details related to Estonian fields of activities and business life-cycle events. Entrepreneurs can use this search to quickly search for any topic of interest.	1,75	Data availability = medium = 2 Data quality = low = 1 Data volume = high = 3 Data preparedness = low = 1	Algorithm simplicity = low = 1	2	Data storage simplicity = medium = 2 Compute simplicity = medium = 2	Estimate hours >= 600 --> high	1,4375
3	Improve DGE process of providing tax and reporting information to entrepreneurs using AI	This use case intends to improve the process of providing most relevant content to the users especially related to taxes and reporting or other relevant factors based on the enterprises' field of activity and other related business events. The aim is to develop a solution which can predict and recommend information to entrepreneurs for the next steps, e.g., event driven report submissions.	1,5	Data availability = medium = 2 Data quality = low = 1 Data volume = medium = 2 Data preparedness = low = 1	Algorithm simplicity = high = 3	2,25	Data storage simplicity = medium = 2 Compute simplicity = high = 3	Estimate hours >= 450	1,5625
5	Guide entrepreneurs through DGE smart suggestions channel related to deadlines and obligations	Enterprises work consistently to meet deadlines and obligations. This use case aims to use an AI solution to enhance the process of suggesting to entrepreneurs how they can move forward with their obligations.	1,75	Data availability = medium = 2 Data quality = low = 1 Data volume = high = 3 Data preparedness = low = 1	Algorithm simplicity = high = 3	2,25	Data storage simplicity = medium = 2 Compute simplicity = high = 3	Estimate hours <= 320 --> medium	2
4	Improve the DGE information channel for entrepreneurs to share and suggest business topic-based information using AI	This use case aims to use AI to solve the problem of finding relevant information from a wide range of topics related to business activities. The idea is to use advanced topic modelling techniques on unstructured data related to tax, obligations, compliance, and regulation etc. Entrepreneurs can subscribe to different topics and receive relevant information.	2	Data availability = high = 3 Data quality = low = 1 Data volume = high = 3 Data preparedness = low = 1	Algorithm simplicity = low = 1	1,5	Data storage simplicity = medium = 2 Compute simplicity = medium = 1	Estimate hours >= 600 --> high	1,375
7	Improve suggestions related to public sector financing opportunities	This use case focuses on improving the process of sharing public sector financing opportunities to companies and entrepreneurs using an AI. The scope is to cover information related to co-financing, grants, loans, innovation funds, guarantees, and other related support offered by public services. Information concerning to procurements opportunities won't be cater in this use case. Furthermore, this use case would be able to take into accounts the information related to enterprises' financial status if available or shared by consent. Such information would enhance the AI-based suggestions for enterprises.	1,75	Data availability = medium = 2 Data quality = low = 1 Data volume = high = 3 Data preparedness = low = 1	Algorithm simplicity = medium = 2	2,25	Data storage simplicity = medium = 2 Compute simplicity = high = 3	Estimate hours <= 320 --> medium	2
2	Enhance DGE process of providing license and compliance information to entrepreneurs using AI	This use case intends to enhance the DGE channel to provide most relevant content to the users especially related to compliance and licenses based on the enterprises' field of activity. Activities depending on the nature of business contains different divisions, and each division consist of some compliance information to operate the business adhering to laws. Such information can be beneficial to business which requires compliance efforts to operate e.g., forestry, mining, manufacturing etc. The aim is to develop an AI-based solution which can take input related to nature of business and recommend information to entrepreneurs for the next steps, e.g., relevant activity and it's divisions which list down related compliance clauses and license required.	1,5	Data availability = medium = 2 Data quality = low = 1 Data volume = medium = 2 Data preparedness = low = 1	Algorithm simplicity = high = 3	3	Data storage simplicity = medium = 2 Compute simplicity = high = 3	Estimate hours >= 450	2

	#	Use Case	Short Description	Data readiness	Technology		Infrastructure		MVP development cost		Ease of implementation
9	10	Improve the process of finding the right public sector experts	This use case aims to enhance through the DGE channel the process of finding the right public sector experts for difficult questions related to business activities. Entrepreneurs can share what they are looking for and an AI can be used to search in unstructured information and suggest the relevant experts and resources.	1,25	Data availability = low = 1 Data quality = low = 1 Data volume = medium = 2 Data preparedness = low = 1	2	Algorithm simplicity = high = 3	2,25	Data storage simplicity = high = 3 Compute simplicity = medium = 2	1,5	Estimate hours >= 450	1,75
10	8	Smart search for finding EU level financing opportunities	This use case aims to improve the process of sharing EU level financing opportunities for Estonian entrepreneurs and enterprises. The focus is to extract information related to grants, funds, loans, and guarantees offered by EU and serve it to the Estonian companies to avail those financing opportunities. Processing information related to EU-level procurements would be out of scope for this use case.	1,25	Data availability = low = 1 Data quality = low = 1 Data volume = medium = 2 Data preparedness = low = 1	2	Algorithm simplicity = medium = 2	2	Data storage simplicity = medium = 2 Compute simplicity = medium = 2	1,5	Estimate hours <= 450	1,6875
11	6	Recommendations related to public sector trainings and events	This use case aims to enhance the process of sharing public sector related events (e.g., conferences, training) and information to entrepreneurs using AI. The AI with textual analysis can be done on the companies, entrepreneurs, and business activities data, and then suggesting relevant public sector events or articles.	1,75	Data availability = low = 2 Data quality = low = 1 Data volume = high = 3 Data preparedness = low = 1	2	Algorithm simplicity = high = 3	2,25	Data storage simplicity = high = 3 Compute simplicity = medium = 2	2	Estimate hours <= 320 --> medium	2
12	9	Public procurement recommendations	This use case intends to help enterprises in the processes related to public procurement and tender submissions using an AI. A solution can use the previous submission history and give recommendations to enterprises to apply and participate in different public procurements.	2	= 2 Data quality = low = 2 Data volume = medium = 2 Data preparedness = low = 2	2	Algorithm simplicity = high = 3	2	Data storage simplicity = medium = 2 Compute simplicity = medium = 2	2	Estimate hours <= 320 --> medium	2
13	12	Smart suggestions to improve business growth	This use case aims to provide support to enterprises who are exporting good and services. The focus is to help enterprises through smart analysis in order that they can enhance their current productivity and market share and expand to their services. The smart suggestions can be related to e.g., highlighting and suggesting that their export related field of activities had growth and in-general there are margins to capture by innovating and expanding business, and also on a high-level what kind of regulations and obligations would be necessary to move forward. The suggestions won't be related to finding grants, loans because there are other use cases and existing services which offer such support to entrepreneurs.	1,5	Data availability = low = 1 Data quality = low = 1 Data volume = medium = 2 Data preparedness = low = 2	2	Algorithm simplicity = medium = 2	2	Data storage simplicity = medium = 2 Compute simplicity = medium = 2	1,5	Estimate hours <= 600 --> medium	1,75
14											

6.3 Annex 2. Survey for the entrepreneurs



Nutikad lahendused ettevõtja digiväravas

Hea ettevõtja!

Majandus- ja Kommunikatsiooniministeerium on koostöös teiste asutustega arendamas **ettevõtjate digiväravat "eesti.ee ettevõtjale"**, mis koondab kokku kõik olulise info ja teenused, mida ettevõtja riigiga asjaajamiseks vajab. Uue välimusega keskkond on ettevõtjatele juba praegu kättesaadav, kuid taustal on käimas mitmed arendusprojektid, et keskkonda tuua järk-järgult uusi teenuseid ja funktsionaalsuseid.

Digivärava arenduses on olulisel kohal ettevõtjate vajadused ja ootused, mida on mitmete varasemate uuringute ja projektide käigus kaardistatud. Seda arvesse võttes on tänaseks formuleeritud 12 tehisintellekti hõlmavat kasutusjuhtu (nutikat lahendust), mis võimaldaksid digiväravas teenuseid pakkuda ja teavet esitleda proaktiivselt ning ettevõtja profiilist lähtuvalt. **Vajame käesolevaga Teie abi, et hinnata, millised nutikad lahendused oleksid Teile kui ettevõtjale antud keskkonnas kõige kasulikumad.**

Tagasiside andmiseks palun hinnake igat võimalikku lahendust individuaalselt skaalal **1-5**. Kui mõni nimetatud lahendustest ei seostu kuidagi Teie ettevõttega, siis palun valige **X**.

Küsimustikule vastamine võtab aega umbes **10 min.**

Vastuse salvestamiseks palun klõpsake kindlasti küsimustiku lõpus olevat "Submit" / "Saada ära" nuppu.

Käesolevat uuringut viib läbi Majandus- ja Kommunikatsiooniministeeriumi analüüsiprojekti partner AS PricewaterhouseCoopers Advisors. Kui Teil on küsimustikuga seoses küsimusi, siis palun pöörduge Elis Raudkett'i poole (elis.raudkett@pwc.com).

Täname kõiki vastajaid!

(not shared) Switch account

* Required

1. Personaliseeritud soovitused vastavalt ettevõtja viimastele tegevustele ettevõtja digiväravas *

Näide: Ettevõtja on digiväravas varasemalt otsinud teavet keskkonnatasude deklareerimise kohta. Seda arvesse võttes, teeb digivärav soovitusi sama või seotud teemaliste teabematerjalide või teenuste kohta.

Palun hinnake kuivõrd kasulikuks peate antud lahendust digiväravas enda ettevõtte jaoks:

- ◯ 1 - ei ole üldse kasulik
- ◯ 2 - pigem ei ole kasulik
- ◯ 3 - nii ja naa
- ◯ 4 - pigem on kasulik
- ◯ 5 - on kindlasti kasulik
- ◯ X - antud lahendus ei ole minu ettevõtte jaoks asjakohane

2. Proaktiivsed soovitused tegevuslubade ning majandustegevusteadete teemal vastavalt ettevõtte tegevusvaldkonnale ning tema äritegevuse käigus esinevatele olukordadele *

Näide: Äsja registreeritud ettevõte planeerib tegeleda kutselise kalapüügiga ning on enda tegevusalaks määranud magevee kalapüügi. Võttes arvesse ettevõtte tegevusvaldkonda, soovitab digivärav taotleda kutselise kalapüügi luba. Ning kuna kalapüügi luba kehtib kuni üks kalendriaasta, annab digivärav loa kehtivusaja lõppedes soovituse luba pikendada, kui ettevõte plaanib samas valdkonnas edasi tegutseda.

Palun hinnake kuivõrd kasulikuks peate antud lahendust digiväravas enda ettevõtte jaoks:

- ◯ 1 - ei ole üldse kasulik
- ◯ 2 - pigem ei ole kasulik
- ◯ 3 - nii ja naa
- ◯ 4 - pigem on kasulik
- ◯ 5 - on kindlasti kasulik
- ◯ X - antud lahendus ei ole minu ettevõtte jaoks asjakohane

3. Proaktiivsed soovitused infoartiklitele ning teenustele maksude ning aruandluse teemal vastavalt ettevõtte tegevusvaldkonnale ning tema äritegevuse käigus esinevatele olukordadele *

Näide: Äsja registreeritud ettevõte registreerib Töötamise registris oma esimese töötaja. Selle tegevuse järgselt soovitab digivärav tutvuda TSD deklaratsooniga ning infomaterjalidega TSD täitmiseks, kui töötajale on välja makstud esimene töötasu.

Palun hinnake kuivõrd kasulikuks peate antud lahendust digiväravas enda ettevõtte jaoks:

- ◯ 1 - ei ole üldse kasulik
- ◯ 2 - pigem ei ole kasulik
- ◯ 3 - nii ja naa
- ◯ 4 - pigem on kasulik
- ◯ 5 - on kindlasti kasulik
- ◯ X - antud lahendus ei ole minu ettevõtte jaoks asjakohane

4. Digiväravas saadaval olevates teemavaldkondades kuvatakse ettevõtjale esmajoones tema ettevõtte jaoks kõige asjakohasemat informatsiooni *

Näide: Restoraniäriga tegelev ettevõte soovib rajada restorani juurde pruulikoja alkoholi tootmiseks ja müümiseks. Ettevõtja digiväravas kuvatakse ettevõtjale huvi pakkuvate teemade seas lähemalt ettevõtte profiilist lähtuvat teavet - antud juhul restoraniäriga seotud vajalikud load, tegevused jmt. Sel moel ei pea ettevõtja kulutama aega, et suurest infohulgast leida tema jaoks kõige asjakohasemat infot. Samuti saab ettevõte omale n-ö tellida (*subscribe*'ida) juurde teavet muu huvipakkuva valdkonna osas - antud juhul nt alkoholi tootmise ja müümisega seotud tingimuste ja kohustuste kohta.

Palun hinnake kuivõrd kasulikuks peate antud lahendust digiväravas enda ettevõtte jaoks:

- ◯ 1 - ei ole üldse kasulik
- ◯ 2 - pigem ei ole kasulik
- ◯ 3 - nii ja naa
- ◯ 4 - pigem on kasulik
- ◯ 5 - on kindlasti kasulik
- ◯ X - antud lahendus ei ole minu ettevõtte jaoks asjakohane

5. Automaatsed teavitused ettevõttele avaliku sektori poolt seatud kohustuste ning tähtaegade kohta *

Näide: Ettevõte saab automaatseid teavitusi regulaarsete ning tema tegevusvaldkonna eripärast tulenevate kohustuste ning tähtaegade osas. Ettevõtja saab valida, milliste kohustuste ning tähtaegade osas ning millise regulaarsusega soovib teavitusi saada.

Palun hinnake kuivõrd kasulikuks peate antud lahendust digiväravas enda ettevõtte jaoks:

- ◯ 1 - ei ole üldse kasulik
- ◯ 2 - pigem ei ole kasulik
- ◯ 3 - nii ja naa
- ◯ 4 - pigem on kasulik
- ◯ 5 - on kindlasti kasulik
- ◯ X - antud lahendus ei ole minu ettevõtte jaoks asjakohane

6. Personaliseeritud soovitused avalike KOV ning riigitasandil koolituste ning seminaride osas ettevõtte profiili arvesse võttes *

Näide: Äsja registreeritud ettevõtte esindaja saab soovitusi osalemiseks erinevatel ettevõtte alustamisega seotud koolitustel ja seminaridel, nt "Alustava ettevõtja baaskoolitus", mida viivad läbi maakondlikud arenduskeskused.

Palun hinnake kuivõrd kasulikuks peate antud lahendust digiväravas enda ettevõtte jaoks:

- ◯ 1 - ei ole üldse kasulik
- ◯ 2 - pigem ei ole kasulik
- ◯ 3 - nii ja naa
- ◯ 4 - pigem on kasulik
- ◯ 5 - on kindlasti kasulik
- ◯ X - antud lahendus ei ole minu ettevõtte jaoks asjakohane

7. Soovitused avaliku sektori poolt pakutavate aktiivsete rahastusvõimaluste (toetused, laenud jt) kohta ettevõtetele vastavalt ettevõtte profiilile *

Näide: Tööstusvaldkonnas tegelevale ettevõttele soovitatakse osaleda EASi ja KredExi ühendasutuse (uue nimega Ettevõtluse ja Innovatsiooni Sihtasutuse) "Tootmisettevõtja ärimudeli muutmise toetuse" taotlusvoorus, kuna tegemist on toetuse sihtgruppi kuuluva ettevõttega.

Palun hinnake kuivõrd kasulikuks peate antud lahendust digiväravas enda ettevõtte jaoks:

- ◯ 1 - ei ole üldse kasulik
- ◯ 2 - pigem ei ole kasulik
- ◯ 3 - nii ja naa
- ◯ 4 - pigem on kasulik
- ◯ 5 - on kindlasti kasulik
- ◯ X - antud lahendus ei ole minu ettevõtte jaoks asjakohane

8. Tark otsingumootor Euroopa Liidu otseste rahastusvõimaluste otsimiseks sobivate märksõnade alusel *

Näide: Sarnaselt Eestis pakutavatele toetustele võimaldab antud lahendus otsida märksõnade abil Euroopa Liidu rahastusvõimalusi. Ettevõtja, kes tegutseb põllumajandusvaldkonnas, võib temale asjakohaste märksõnade abil leida seotud toetuseid, laene jt rahastusvõimalusi.

Palun hinnake kuivõrd kasulikuks peate antud lahendust digiväravas enda ettevõtte jaoks:

- ◯ 1 - ei ole üldse kasulik
- ◯ 2 - pigem ei ole kasulik
- ◯ 3 - nii ja naa
- ◯ 4 - pigem on kasulik
- ◯ 5 - on kindlasti kasulik
- ◯ X - antud lahendus ei ole minu ettevõtte jaoks asjakohane

9. Teavitused avaldatud riigihangete kohta ettevõtetele, kes on varasemalt sarnastel hangetel osalenud või hanke pakkumuses olevad andmed seostuvad ettevõtte tegevusalaga *

Näide: Ettevõte, kelle tegevusvaldkonnaks on ehitus, saab digiväravast soovitusi ehitusega seotud tegevusvaldkondade hangete kohta.

Palun hinnake kuivõrd kasulikuks peate antud lahendust digiväravas enda ettevõtte jaoks:

- ◯ 1 - ei ole üldse kasulik
- ◯ 2 - pigem ei ole kasulik
- ◯ 3 - nii ja naa
- ◯ 4 - pigem on kasulik
- ◯ 5 - on kindlasti kasulik
- ◯ X - antud lahendus ei ole minu ettevõtte jaoks asjakohane

10. Sobiva riigiasutuse ametniku poole suunamine ettevõtja poolt kirja pandud küsimuse või märksõnade alusel, mille kohta info digiväravas puudub või on puudulik *

Näide: Ettevõtja soovib leida kõige asjakohasemat kontakti leiutisele kaitse saamise teemal. Ettevõtja kirjutab kontakti otsimiseks vastavasse otsingumootorisse märksõnad "leiutis", "kaitse" ning digivärav pakub antud teemale sobivaks Patendiameti Patendiosakonna (võimalusel konkreetse ametniku) kontaktandmed.

Palun hinnake kuivõrd kasulikuks peate antud lahendust digiväravas enda ettevõtte jaoks:

- ◯ 1 - ei ole üldse kasulik
- ◯ 2 - pigem ei ole kasulik
- ◯ 3 - nii ja naa
- ◯ 4 - pigem on kasulik
- ◯ 5 - on kindlasti kasulik
- ◯ X - antud lahendus ei ole minu ettevõtte jaoks asjakohane

11. Tark otsingumootor, mis võimaldab ilma täpset märksõna teadmata leida asjakohaseid tulemusi *

Näide: Ettevõtja soovib saada infot toidu käitlemisega seotud kohustuste ja teenuste kohta. Ettevõtja kirjutab otsingumootorisse nt ühe järgnevatest märksõnadest: "toit", "toitlustus", "toitlustamine", "mesi", "munad", ning saab samaväärse ning asjakohase info otsingu tulemusena.

Palun hinnake kuivõrd kasulikuks peate antud lahendust digiväravas enda ettevõtte jaoks:

- ◯ 1 - ei ole üldse kasulik
- ◯ 2 - pigem ei ole kasulik
- ◯ 3 - nii ja naa
- ◯ 4 - pigem on kasulik
- ◯ 5 - on kindlasti kasulik
- ◯ X - antud lahendus ei ole minu ettevõtte jaoks asjakohane

12. Soovitused ekspordiga alustamiseks või laienemiseks uutele turgudele *

Näide: Ettevõte, kes juba ekspordib või planeerib eksportida, saab vastavalt tema tegevusvaldkonnale lähemalt tutvuda, millistel turgudel oleks tema ettevõttel kasvupotentsiaali.

Palun hinnake kuivõrd kasulikuks peate antud lahendust digiväravas enda ettevõtte jaoks:

- ◯ 1 - ei ole üldse kasulik
- ◯ 2 - pigem ei ole kasulik
- ◯ 3 - nii ja naa
- ◯ 4 - pigem on kasulik
- ◯ 5 - on kindlasti kasulik
- ◯ X - antud lahendus ei ole minu ettevõtte jaoks asjakohane

Ettevõtte tegevusvaldkond *

- [ ] Põllumajandus, metsamajandus ja kalapüük
- [ ] Mäetööstus
- [ ] Töötlev tööstus
- [ ] Elektrienergia, gaasi, auru ja konditsioneeritud õhuga varustamine
- [ ] Veevarustus; kanalisatsioon, jäätme- ja saastekäitlus
- [ ] Ehitus
- [ ] Hulgi- ja jaekaubandus; mootorsõidukite ja mootorrataste remont
- [ ] Veondus ja laondus
- [ ] Majutus ja toitlustus
- [ ] Info ja side
- [ ] Finants- ja kindlustustegevus
- [ ] Kinnisvaraalane tegevus
- [ ] Kutse-, teadus- ja tehnikaalane tegevus
- [ ] Haldus- ja abitegevused
- [ ] Avalik haldus ja riigikaitse; kohustuslik sotsiaalkindlustus
- [ ] Haridus
- [ ] Tervishoid ja sotsiaalhoolekanne
- [ ] Kunst, meelelahutus ja vaba aeg
- [ ] Muud teenindavad tegevused
- [ ] Kodumajapidamiste kui tööandjate tegevus; kodumajapidamiste oma tarbeks mõeldud eristamata kaupade tootmine ja teenuste osutamine
- [ ] Eksterritoriaalsete organisatsioonide ja üksuste tegevus
- [ ] Tegevusala määramata
- [ ] Other:

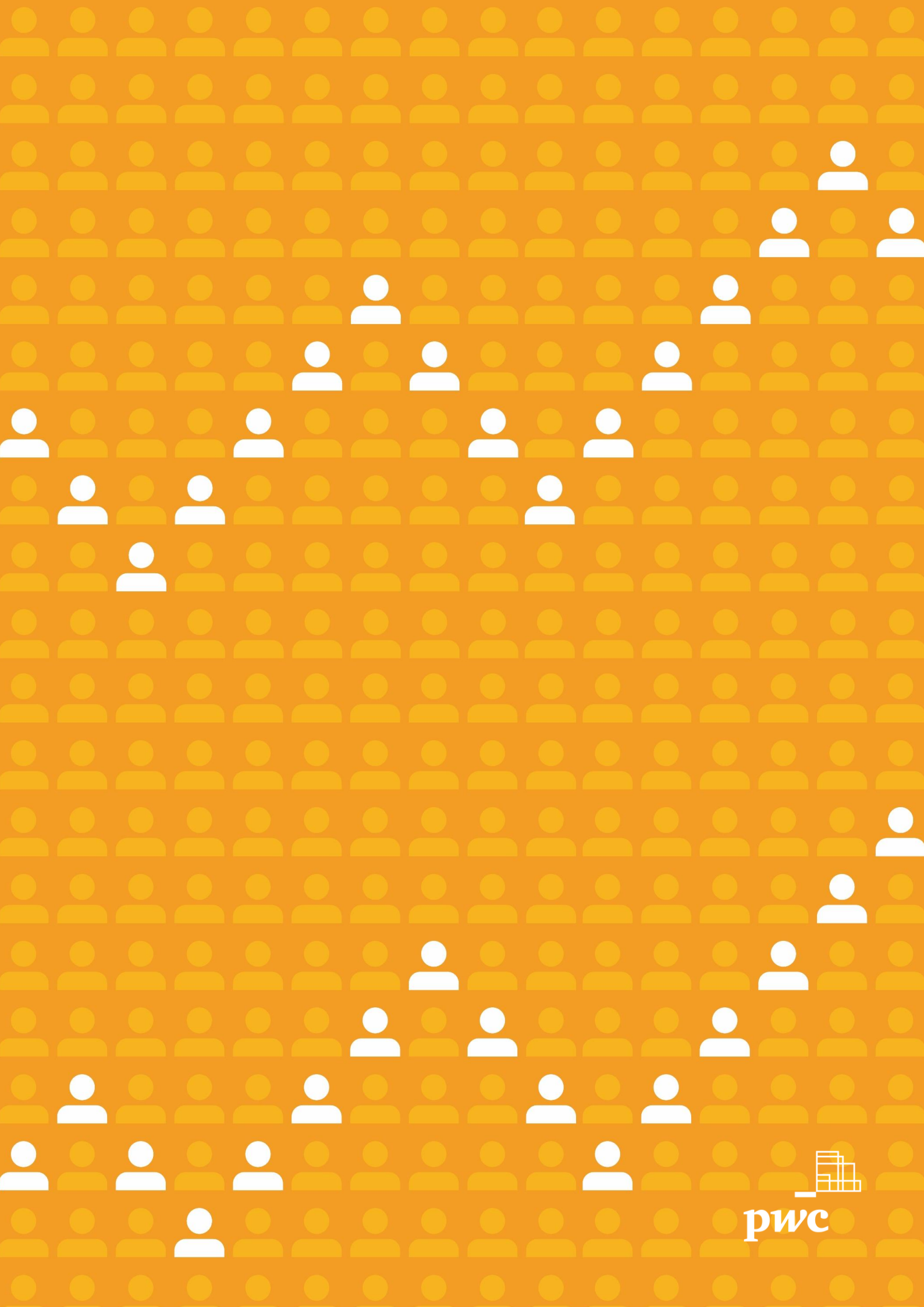
Ettevõtte töötajate arv *

- ◯ 0
- ◯ < 10
- ◯ < 50
- ◯ < 250
- ◯ üle 250

Soovi korral võite lisada oma mõtteid või kommentaare kirjeldatud lahendustega seonduvalt

Your answer

© 2023 AS PricewaterhouseCoopers Advisors. Kõik autoriõigused kaitstud. "PwC" viitab Eesti seaduste alusel asutatud ja tegutsevale äriühingule AS PricewaterhouseCoopers Advisors või kontekstist lähtuvalt teistele PricewaterhouseCoopers International Limited rahvusvahelise võrgustiku liikmetele, kellest igaüks on eraldiseisev ja sõltumatu juriidiline isik.


pwc