European Commission

JRC SCIENCE FOR POLICY REPORT

AI Watch
European Landscape on the Use of Artificial Intelligence by the Public Sector

Annex II

Case studies description

Joint Research Centre

EUR 31088 EN

Contact information
Name: Francesco Pignatelli
Address: Via Enrico Fermi 2749, 21027 Ispra (VA), Italy
Email: EC-AI-WATCH@ec.europe.eu

EU Science Hub
https://ec.europa.eu/jrc
JRC129301
EUR 31088 EN
Luxembourg: Publications Office of the European Union, 2022

EN PDF ISBN 978-92-76-53435-8 ISSN 1831-9424 doi:10.2760/481674

How to cite this report: Tangi L., van Noordt C., Combetto M., Gattwinkel D., Pignatelli F., AI Watch. European Landscape on the Use of Artificial Intelligence by the Public Sector. Annex II. Case studies description. EUR 31088 EN, Publications Office of the European Union, Luxembourg, 2022, ISBN 978-92-76-53435-8, doi:10.2760/481674, JRC129301.

Intelligent Control Platform

Danish Business Authority

Function of Government	General public services
Status	In production
AI Classification	Machine Learning, Automated Reasoning
Interaction	Government 2 Government
Main value driver	Efficiency

The organisation

The Danish Business Authority (DBA) endeavours to create predictable and responsible business conditions, to make it as easy as possible to run a business in Denmark. It has two main goals: (i) ensure the proper conditions for businesses to thrive and to start a business and (ii) create a level playing field for all businesses.

The solution

The intelligent control platform (ICP) is a digital platform that among its main tasks, is providing an automated assessment of the risk of a company being fraudulent. The system calculates a risk score following a variety of factors from the data collected at the time of a company registration. The platform is capable to predict fraud with 75% accuracy based on bankruptcy and VAT fraud. The platform serves model output directly to case handling systems Among these are:

1) Identity (it controls all identity papers for foreign business actors)
2) Assets (it controls that fictional assets are not inserted)
3) Audits ('Weaponised' unstructured data concerning negligence)
4) 1st line (Control newly registered business for concerns of fraud)

Technical information

The platform uses predictive machine learning (ML) trained with Graph Database with data collected from over 800.000 companies, plus a variety of other data coming from administrative databases merged them in the Graph database. This enables to use ML with graph analysis as well as create ML models from complex datasets. Data is analysed using various ML models (about 30 models) in combination by operators through the platform interface.

Drivers and challenges

- The management was supporting the development of the system thanks also to high attention from the team on managerial involvement and awareness
- Enthusiasm from the data science team and general trust in their work were main drivers for the development
- A lot of effort has been put on communication activities to ensure that the data science unit understands the needs from the operational side and vice versa
- The XRAI Methodology has been used to ensure that a precise description of the needs and expectations of the business is available. This developed methodology includes questions asked to domain experts for model development. The output of this process is a file is a machine readable file which will be stored together with the source code with that individual model and exposed on an endpoint by the model in production
- The development of the project followed a larger discussion on the organisational strategy on working with AI. In particular data scientists were hired for guaranteeing internal capacity and avoiding full reliance on third parties

Intelligent Control Platform

Danish Business Authority

Risks and mitigation measures

- Domain users were involved since the beginning for avoiding possible misinterpretations. In fact, technical employees are not capable of understanding the nuances and interpretations of the domain area
- Strong effort and time dedicated to the maintenance and update of the models
- AI system takes and transforms data, hence the data science unit decided to take full responsibility and full ownership of the data used, even though they initially come from different departments/units within the Business Authority or even from third parties
- A framework, named Record Keeper, has been built to ensure the transparent, responsible, and accurate use of AI. This is a monitoring log which observes all actions taken regarding data handling and decision. Events are stored in a DAG that allows explainability through giving a causal path of the different events which led to the basis for a decision. Furthermore, the Record Keeper allows more awareness of potential breaches of security.

Curiosity box

Managerial trust and support is something that can be created and developed over time. This was the case of the Danish Business Authority that, before starting with the project, created a 'fake' system, using random data with the sole purpose of making the project explainable to the management and arise trust and support. Thanks to this approach data scientists were able to open the black box of algorithms. For example, they clarified all privacy and ethical concerns and show the high precision rate of the system. Moreover, they showed the final interface that is the result of the analysis but also the origin of those data. This made the management realise the ethical trade-offs associated with the use of AI but also the potential benefits and the need for managerial commitment. As a result, management is now on board with a supportive attitude but also with an aware approach that led them to make conscious decisions.

Useful links

- GraphTalk Copenhagen - Fraud Detection with Graphs (slideshare.net)
- Fraud Detection with Graphs at Danish Business Authority - YouTube

Contact details

Marius Hartmann
MarHar@erst.dk

Eva, targeted COVID-19 border checking

Greek National Government

Function of Government	General public services
Status	Not in use anymore
AI Classification	Machine Learning, Planning and Scheduling
Interaction	Government 2 Citizen
Main value driver	Efficiency

The organisation

Within the Greek national Government, several ministries have been involved, above all the Ministry of Digital Governance, the Ministry of Civil Protection, the Ministry of health and social solidarity. The solution has been developed in close cooperation with the USC Marshall School of Business, and the Wharton School of Business at the University of Pennsylvania

The solution

Between August and November 2020, in the midst COVID-19 crisis, the Greek Government has been trialling an AI system in border control points which helped with the selection of the travellers to test upon arrival. The purpose of the system, named Eva, was to effectively allocate the scarce PCR tests within the crowded summer tourism season of Greece. The system analyses data collected from the Passenger Locator Form that people had to fill in advance of their arrival to Greece. Eva allowed the identification of 1.25x- 1.45x more infections than a randomised testing protocol. Currently, EVA is not in use anymore: after the summer peak, fewer tourists were coming so it was possible to either test everyone upon arrival or to close the borders for non-essential travel. At the beginning of the summer season in 2021, vaccines were already introduced, and the availability and PCR tests were no longer scarce hanging the overall strategy of Greece.

Technical information

The system has been developed in a few months, given the emergency. Hence, it was designed in such a way as to be practically useful, effective, and deployed quickly. First, travellers must complete a passenger locator form at least 24 h before arrival, containing (among other data) information on their origin country, demographics, and point and date of entry. Second, Eva estimates the prevalence of SARS-CoV-2 among asymptomatic travellers. The estimation is based on various machine learning algorithms (reinforcement learning, regression, empirical Bayes and multi-armed bandit algorithms). Third, EVA targeted a subset of travellers for allocating the scarce tests. Fourth, Eva's prevalence estimates were used to recommend risky countries; for those countries specific restrictive measures were taken.

Drivers and challenges

- The high complexity of the project is related to the number of different actors involved, with completely different backgrounds, from data scientists to epidemiologists and all the logistics necessary for transporting the tests
- The involvement various stakeholders (scientific and political) was important in pushing down decisions and championing the system internally
- One of the main factors which made this project a success was the general awareness that the use of AI was not solely a technical problem, but an operational and logistic challenge as well

Eva, targeted COVID-19 border checking

Greek National Government

Risks and mitigation measures

- During the design of the Passenger Locator Forms, discussions with lawyers were held to ensure legal and ethical compliance. Trade-offs showed up between privacy and the capability in assessing individual risks. For instance, the information on the occupation was excluded as, even though it would have been extremely useful, it was considered too invasive
- The information collected was pseudonymised and only accessible in an aggregated pseudonymised way. At the same time, the contact tracing teams had to be able to access the previously anonymised data, hence, the database was designed with strict access rules

Curiosity box

Some blogs online were spreading misinformation to travellers: they suggested declaring Athens as the city of birth for avoiding to be tested. Following this – false - rumour, a lot of travellers from Athens came up in the data despite they were not Greek – Athens from Spain, France and other impossible combinations. Some people tried to game AI systems which could lead to unexpected data inputs following its deployment. As such, reviewing the data and the performance of the AI system was crucial to avoid any irregularities and to take action if appropriate. In addition, the system had some automatic flags in case of presumed irregularities to be discussed with the experts as to understand why this occurred – and if any actions are to be taken after.

Useful links

- An algorithm may have helped slow the spread of COVID-19 in Greece (usc.edu)
- An AI-based tool "helped Greece reopen its borders" – AlgorithmWatch
- We Helped Greece Build an AI System to Make Covid-19 Testing More Efficient. Here's What We Learned. (entrepreneur.com)
- Efficient and targeted COVID-19 border testing via reinforcement learning - PubMed (nih.gov)
- This algorithm screened travelers to Greece for COVID-19 • The Register
- Efficient and targeted COVID-19 border testing via reinforcement learning | Nature
- Dean's Dialogue: How Algorithms Helped Greece Cope with Covid-19 on Vimeo
- Could AI Help Slow the Spread of the Coronavirus? (fairobserver.com)
- Interpretable OR for High-Stakes Decisions: Designing the Greek COVID-19 Testing System by Hamsa Bastani, Kimon Drakopoulos, Vishal Gupta

Contact details

Vishal Gupta
guptavis@usc.edu

Kimon Drakopoulos
drakopou@marshall.usc.edu

Hamsa Bastani
hamsab@wharton.upenn.edu

Reducing night noise through nudging

Municipality of Leuven

Function of Government	General public services
Status	In production
AI Classification	Audio Processing, Machine Learning
Interaction	Government 2 Citizen
Main value driver	Efficiency

The organisation

Leuven is the capital and largest city of the province of Flemish Brabant in the Flemish Region of Belgium. It is one of the largest municipalities in Belgium with approximately 100.000 inhabitants. Leuven is home to the Katholieke Universiteit Leuven, the largest and oldest university of Belgium and the oldest catholic university still in existence. Hence, the city has a large student population.

The solution

In the city of Leuven, streets are often crowded at night as many people walk there following a night out. This regularly causes too much noise at night for the residents. For solving the issue sound meters have been installed and an application for citizens reporting has been developed. The sound meters will be capable of detecting the types of noise using Artificial Intelligence, to also better understand what kind of night noises are disturbing the local residents. The project is now in the pilot phase in one of the most crowded streets of the city (Naamsestraat). In a later phase of the project, there is the aim to use nudging techniques to intervene automatically in case noise is detected or prevented. Two scenarios will be experimented: dim the public lights from 50% to 30% capacity (= calming effect) and intensify the lights from 50% to 100% capacity (= shock effect). Any other forms of interventions, such as projecting messages on the street, will be explored and discussed.

Technical information

The noise meters (7 now in the piloting phase) are very accurate in measuring objective parameters of noise such as decibel and pitch. Moreover, they have smart modules installed capable of recognising three main categories of noise: traffic-related noise, music and human sounds. The solely objective data has been not considered sufficient for the project as there is a strong need to understand why and from which sounds people are getting disturbed. Hence residents could install an app, and if they get disturbed at night due to the noise, declare which kind of noise disturbed them and how long the noise lasted. These data will be used to train the whole system.

Drivers and challenges

- The Flemish Regional Government gave funds for IoT projects that covered the 80% of the total costs for piloting. For the upscaling the city will have to cover the 50% of the costs, even though whether these funds will be available in the future is still unclear
- The project has been assisted by using innovative procurement to find an external party that could assist with the instalment of the noise sensors and develop the app. Using external parties is in fact, obligatory as part of the funding requirements
- The municipality of Leuven is cooperating with the municipality of Antwerp, which is doing a similar project with the sound meters and the app

Reducing night noise through nudging

Municipality of Leuven

- Involving citizens is extremely challenging. Some have expressed skepticism, as the same people have been calling the police for numerous years – in their perspective to no avail
- At least 250 responses are needed at least to have some form of recognition possible. This is not an easy target, given that the pilot is just in one street with not a lot of people living there

Risks and mitigation measures

- Data from the app are collected anonymously. Due to this requirement, it is challenging to see who is providing the notifications on the app and whether it is coming from only one, or multiple citizens
- The noise meters could listen to people's conversations. However, after a check with the legal office, this was forbidden. Short audio samples mixed with anonymised data would have been extremely useful to train the systems, but this was not in line with the privacy legislation
- A list of principles regarding interoperability, privacy and security have been defined to assess the system. This checklist has been developed and used in cooperation with all the smart city projects
- Renting the noise meters from external suppliers has been seen also as a measure for mitigating risks: any malfunctioning of the system would be part of the contractor's contractual responsibility

Curiosity box

For the municipality of Leuven measuring the noise is not a new intervention. By local regulation, every cafe in Leuven must install a noise meter because playing music louder than 90 decibels is not allowed. Hence the city uses a platform to collect and analyse all these noise meter data. This monitoring is done by the environmental department. The synergies with this project allow a big advantage for the municipality. First, a colleague from the environmental department was involved in working on interpreting the noise data and creating an index to visualise noise information. Second, the supplier providing the two platforms (for café and street noise data) is the same, this will allow easier integration of the two sets of data. In fact, now the municipality is also working on the creation of a unique Smart City Data Platform

Useful links

- Nachtlawaai verminderen met technologie | Stad Leuven
- Infosessie bewoners - Nachtlawaai verminderen - YouTube

Contact details

Helena Schulpé
Helena.Schulpe@leuven.be
smartcity@leuven.be

Unlocking digitised documents and correcting OCR

Luxembourg National Library

Function of Government	General public services
Status	Implemented
AI Classification	Computer Vision, Natural Language Processing
Interaction	Government 2 Citizen
Main value driver	Public Access

The organisation

Bibliothèque nationale du Luxembourg
Nationalbibliothéik

The National Library (BnL) is the country's main heritage, scientific and research library. In addition to more than 1.8 million paper publications, it is offering its users an increasing number of documents in digital format: e-journals, e-books and databases. Much more than a place of study and documentation, the BnL is also a cultural meeting place. It regularly organises conferences, exhibitions and events on a wide range of topics.

The solution

Since 2003 the BnL started a process of digitisation of historical newspapers and books. At the moment, the library has digitised already more than 900.000 pages of national newspapers. Visitors as well as researchers at the library can use the files freely. For doing this, BnL over the years relied on many different suppliers (a different one every one or two years). This choice led to different and uneven results, considering also that the documents are in several languages, which have evolved over the years, and this is often difficult to detect. Moreover, over the years, the quality of the Optical Character Recognition (OCR) software evolved, hence some documents are now considered low quality due to poor image scanning or poor quality of the software. These documents are difficult to find and use for researchers. To solve this BnL developed an AI system that operates on top of the results of the different OCR for correcting them. About two-thirds of the original results have been improved through this AI.

Technical information

The AI system is a collection of different tools and components, it has been developed fully in-house by the library, using open-source solutions.

1. One component is a tool that can recognise the language in the existing OCR transcription. Luxembourg is a multilingual country, so there is historical content available in German, French, Luxembourgish, English and sometimes other languages
2. Another component is used to determine the current quality of the OCR by looking at the transcribed text. The tool uses the K-nearest neighbours technique. The tool decides whether the current OCR is acceptable or not
3. Finally, for non-acceptable documents, an in-house OCR is run for adjusting the digital copy

Drivers and challenges

- The project follows a Proof of Concept done a few years earlier. The project wasn't scaled up at that time due to a lack of resources only recently made available: an additional IT expert was hired
- The past digitisation effort was key: it allows a simpler software selection as well as an easy way to select manually corrected ground truth to train the system

Unlocking digitised documents and correcting OCR

Luxembourg National Library

- Funding has been made available through the broader eLuxemburgensia meets AI (eLmA) project, partially funded through the national AI4GOV initiative, to further assist in the enrichment of the digitised collections of the library
- The hiring process was considered the most difficult obstacle to overcome: it was challenging to find someone who is qualified and willing to work for the organisation

Risks and mitigation measures

- The involvement of the Heritage Department was important to ensure the correctness of the transcription of historical texts
- Risks are considered rather minimal in this case, as there is limited risk to individuals or the reputation involved. There might be some GDPR-compliance issues in the future, but none were encountered yet

Curiosity box

The whole eLmA project also includes two additional projects, one on Entity Recognition and the other on Entity Linking. Entity Recognition allows the system to identify named entities, such as names, place names as well as other elements which have a definite name. This allows for searching and browsing for specific names. However, due to the long period (200 years) of the newspapers present in the database, Entity Recognition is difficult as named entities could be linked to a wide variety of different topics (for example the word 'Napoleon' might refer to the political French leader, but also to a street, a square, a project, etc.). Entity Linking then enables the linking of these entities with other libraries using unique identifiers. This works on top of the named entity recognition. These other projects are now in a more embryonic phase and have not been applied (yet), also due mainly to the accuracy issues still to solve.

Useful links

- Optimising open data from Luxembourg's historical newspapers | Joinup (europa.eu)
- OCR improvements through machine learning methods and the impact on the long term preservation of digitized content - Open Preservation Foundation
- The AI4Gov initiative - government.lu (gouvernement.lu)
- [2110.01661] Rerunning OCR: A Machine Learning Approach to Quality
- Assessment and Enhancement Prediction (arxiv.org)

Contact details

Yves Maurer
yves.maurer@bnl.etat.lu
eluxemburgensia@bnl.etat.lu

Object Detection Kit

City of Amsterdam

Function of Government	General public services
Status	Not in use anymore
AI Classification	Computer Vision, Machine Learning
Interaction	Government 2 Government
Main value driver	Efficiency

The organisation

City of Amsterdam

Amsterdam is the capital and most populous city of the Netherlands; with a population of more than 800k inhabitants. The organisation of the municipal government is complex: overall the local government counts more than 18.000 employees with dozens of different departments. The municipality has significant expertise in AI and ICT, with about 1200 different people working on ICT and about 300-400 people on data-related tasks. Moreover, there are about 80 data scientists.

The solution

The city of Amsterdam has used the Object Detection Kit to detect various urban challenges. In this first run, the project focused on detecting garbage. Although different elements then garbage have been explored, they have never been implemented this into the system.This is done by analysing imagery collected from the pictures taken by smartphones installed into vehicles driving around in the city. The AI application automatically identifies trash on the street and shares this with the garbage management services of the city to act and solve the issue. After the pilot, the project was discontinued – or on hold for now for several reasons, from funding issues to the need for more organisational commitment and the need to resolve large ethical and legal concerns.

Technical information

The Object Detection Kit is an image recognition system that can detect what type of garbage is available in the streets. It is based on public computer vision models which are available on the web. The development itself took about two years, with one year of intensive development and testing. The data for training the system was collected by a person already tasked with going around the city to see what is going on, as a neighbourhood watch. This person was asked to drive around a few times a week for a period of 3-4 months to collect the data. Currently, there is a large, annotated dataset of 25.000 images that have been used for training the system.

Drivers and challenges

- The project was focused on making the technology available for a low budget, which is why smartphones have been used for taking pictures, instead of more innovative - and expensive - 360 degrees cameras. Funding became available with a subsidy from the Dutch Government.
- The pre-existing technical expertise of the city was fundamental, as the city was not relying on any external supplier
- There was important support from the people collecting garbage on the street. People that had to work from home due to the COVID-19 situation were dedicated to train the system, annotating and categorising the images
- Managerial and political commitment is key, especially for moving beyond pilots, and was one of the reasons for discontinuing the project

Object Detection Kit

City of Amsterdam

Risks and mitigation measures

- There is a strong emphasis on transparency: detailed documentation on the AI system is available online
- Images collected on the street have several ethical and privacy concerns to be taken into account. The system was programmed to delete images or at least sensitive information after detecting the garbage
- Moreover, cameras already installed for other purposes were not used for ethical and legal issues (for example there were cameras installed on cars for parking enforcement purposes)

Curiosity box

Innovation is key for the city of Amsterdam. In fact, the city in the last year created an innovation team that develop and test innovative projects. Approximately 80 people working on Data Science for the City of Amsterdam, the innovation team has about 10. The team work without the pressure for the final implementation of projects: it is worth having a lot of projects that fail or without the possibility for their production if few of them are successful. This team is one of the flagships of the city, characterised by an innovative culture and modus operandi for the public sector. For example, the working language is English, which allows attracting also non-native Dutch speakers. Moreover, computing power and last-generation equipment are guaranteed to all employees. The team works closely with the university network, to test and research new application, just for the experimental sake of it. The team is the most popular organisation in the NL for master thesis students, every year they support around 200 theses and then they hire around 25 of those young experts. In doing this, the city take important steps into the direction of attracting young talents.

Useful links

- New digital system will keep Amsterdam's streets cleaner and safer | I amsterdam
- ABS-onderzoek houdt Amsterdam schoon: real-time beeldherkenning voor verzamelen zwerfafval - Universiteit van Amsterdam (uva.nl)
- Using AI to Keep City Clean Makes Amsterdam 2021 GO SMART Award Winner - EE Times Asia (eetasia.com)
- Object Detection Kit - openresearch.Amsterdam
- ODK.ai

Contact details

Maarten Sukel
m.sukel@amsterdam.nl

OTT - decision-support tool for consultants

Estonian Unemployment Insurance Fund (Eesti Töötukassa)

Function of Government	Social Protection
Status	In production
AI Classification	Machine Learning. Automated Reasoning
Interaction	Government 2 Government
Main value driver	Efficiency

The organisation

Eesti Töötukassa

Eesti Töötukassa administers unemployment insurance benefits. It is a quasi-governmental organisation, that performs its activities independently, but on the basis of a mission and of operational rules defined by law. Its key responsibilities are: (i) provision of Active Labour Market Measures, (ii) payment of unemployment and work ability benefits and compensations, (iii) Work capability assessment. Eesti Töötukassa has a two-level structure with one head office and 15 regional offices

The solution

OTT (this is how the system has been named) is an AI system used in the Estonian Unemployment Insurance Fund which aims to assist its consultants with providing insights predicting the chances of their client – an unemployed person - getting a new job. It calculates the probability of moving into employment for unemployed persons and the probability of becoming unemployed again for people that got a new job. The system also presents the factors affecting these probabilities. Following this assessment, consultants have all the needed information for prioritising certain clients and for tailored supporting activities. This support was also necessary due to the high staff turnover and the lack of senior consultants, ensuring to the client the same treatment independently from the experience of the consultant.

Technical information

OTT uses a random forest machine learning model. The peculiarity of the system is the integration of a large amount of variables (over 40) that tried to capture also soft data sources which were previously very challenging to capture or to include in official registries, such as behavioural factors. For example, the system collects the information related to how the client has registered to the service; whether online or coming to the counter. Such an indicator is used as one of the proxies for measuring digital skills. The system has been trained and tested based on the last five years' unemployment data. The scoring of the clients is recalculated every night. Moreover the model is retrained every quarter in a semi-automatic way: based on the input of the system consultants meet for discussing and modifying the algorithm. This was an explicit choice of the organisation that recognise the central and necessary role of the consultants, that cannot be completely overcome by automatic decision-making.

Drivers and challenges

- The development is the result of a partnership among various organisations: (i) the University of Tartu with a data science team is in charge of maintaining the data, the model and retraining it; (ii) a software company is responsible for the integration of the system into the existing ICT systems; (iii) another partner is in charge of the building and maintenance of the data lake; (iv) lastly, another contractor provides the hosting and the computer power to train the models

OTT - decision-support tool for consultants

Estonian Unemployment Insurance Fund (Eesti Töötukassa)

- Medium and senior management are extremely committed to the project: they see the benefit for the consultants but also the benefit for their coordinating role as they could reassign clients to a different consultant based on the information available (i.e. the similarity of the profiles, equilibrium of labour force given that high-risk client takes more time, etc.)
- Responses by the consultants working with Ott vary, while some of them are appreciating and heavily using the system others are more sceptical and not fully aware of the potentialities of the tool. In particular, in smaller regional offices, personal knowledge from the consultants of the clients –often due to personal connections - is overcoming the system score, while in larger cities, with a higher volume of clients, the use of Ott is seen as more useful
- The organisation encountered the need for training staff to gain a very general level of understanding of machine learning to understand the capabilities of the system and to have a critical approach toward the final score. This type of awareness of the employees is and will be crucial for a sustainable usage of OTT
- Atypical cases are challenging: the system often misclassified them. Those cases need to be carefully assessed by the consultants
- Funding to develop Ott came from the IT budget of the Estonian Unemployment Insurance Fund

Risks and mitigation measures

- The human competences have been never overcome by the system. In fact the score is just a support and the final choice of prioritisation is the full responsibility of the consultant
- On the same line of thoughts, the model is retrained with the involvement of the team, which put its expertise in the refinement of the algorithm

Curiosity box

The personal motivation of clients is heavily influencing the success and/or failure of the services offered by the Insurance Fund. In fact, motivation is one of the main drivers that affect the possibility to get a job. Motivational elements are affecting also OTT: a person with high (or low) motivation can 'break' the model, making the probability of unemployment less predictable. An internal discussion has been conducted on the possibility of sharing the score calculated by OTT with the clients for being as transparent as possible. However, there was the concern that someone with a high-risk score may demotivate or alternatively if it is a low-risk client, the person may not be enough committed assuming they will easily find a job. In the end, the Insurance Fund decided, for now, to don't share the information, considering the demotivation a serious risk to avoid.

Useful links

- OTT – an AI-powered success story in the public sector - Nortal
- AI to help serve the Estonian unemployed - e-Estonia
- Our app won first prize – ECePS – the ERA Chair in e-Governance and Digital Public Services (ut.ee)
- How Estonia is using AI to tackle unemployment - Tech Monitor

Contact details

Center of IT Impact Studies
citis@ut.ee

Automation of subtitling videos and audios

Finnish Tax Administration

Function of Government	General public services
Status	In production
AI Classification	Audio Processing, Machine Learning
Interaction	Government 2 Government
Main value driver	Personalisation

The organisation

The Finnish Tax Administration is the revenue service of Finland. It is a government agency steered by the Ministry of Finance. The Tax Administration collects taxes to safeguard the functioning of the Finnish society through credible tax control and by providing proactive guidance and good service. The goal is to help all taxpayers conduct their tax matters independently and correctly Every year, tax recipients receive over €50 billion in tax revenue.

The solution

The Finnish Tax Administration regularly publishes videos with important information to citizens. These videos need to be accessible by everyone, moreover the directive (EU) 2016/2102 is requiring high accessibility standards. The AI system is based on understanding speech and transforming it into text. It is used to provide subtitles on all the videos and is part of a wider initiative within the administration to use Speech-to-Text technologies in various use cases. At the moment of writing, the AI solution has been implemented and it is in use, after passing an internal security audit.

Technical information

The transcription service is based on three different components. The first part is a portal that acts as an interface: employees can place the videos, which will then trigger the transcription service. This part has been developed internally. The second part is the speech to text technology itself, which uses Microsoft Azure's Cognitive Services as the transcription element. The third part is software which (i) handles the subtitling and the formatting and (ii) controls and checks the transcription. This is based on Python and has been developed internally by the agency. In the end, the user will receive an email with the transcription in a .srt file (commonly used for subtitling). The user can edit the file if any mistakes are detected.

Drivers and challenges

- The Finnish Tax Administration has a long history working with Microsoft services and products. This eased the selection of the supplier and the development of the project
- There was the need of adjusting the results obtained through Microsoft Azure, hence one in-house developer was tasked mainly in developing the Python layer

Risks, Impacts and Reuse

- The Finnish Tax Administration has defined a list of ethical principles to be checked before the usage of an AI system
- The project has been considered (very) low risk since no personal or sensitive data was used within this AI system or is supposed to be processed
- For ensuring high quality the administration is exploring the possibility to have a set of videos processed once a year, to see if the results are changed for better or for worse

Automation of subtitling videos and audios

Finnish Tax Administration

Curiosity box

The project started with a completely different goal: to record and transcribe incoming phone calls from customers and use the text for quality control. However, this initiative ran into various implementation challenges which led to its termination, at least for now. For instance, there were various privacy-related challenges as citizens' phone calls often involved sensitive or even private information. Moreover, the cost was to high: most speech to text translation services charge by the minute, and the Tax Authority receives about 2 million phone calls per year. Maybe in the future, the technology will be more accessible and the original project will be brought back to light.

Useful links

- Finnish Tax Administration's ethical principles for AI – vero.fi

Contact details

Laakso Mikko
mikko.laakso@vero.fi

Estimation of income for those paying by modules

Spanish Tax Agency

Function of Government	Social Protection
Status	Proof of Concept
AI Classification	Automated Reasoning, Optimisation
Interaction	Government 2 Government
Main value driver	Efficiency

The organisation

The Tax Agency is an entity of public law controlled by the Ministry of Finance and Civil Service. Its mission is to promote compliance by citizens with their tax obligations. To this end, it develops two lines of action: on the one hand, the provision of information and assistance services to taxpayers to minimise the indirect costs associated with compliance with tax obligations and, on the other hand, the detection and regularisation of tax breaches through control actions.

The solution

In the Spanish Tax Authority, an AI system is being used which estimates the income of Small and Medium Enterprises (SMEs) as well as of self-employed individuals who have decided to pay their taxes by module rather than defining an exact amount of income. For certain activities, self-employed and SMEs can opt for the reporting of objective measures of their economic activity, these measures are called "modules". The modules are used by the Tax Authority to calculate the estimated amount of their income. To generate more accurate statistics and forecasts of taxes income, the Statistics unit of the Tax Agency developed a model based on similar taxpayers (in terms of activity, size of the establishment, location, etc.) who operate with similar income/taxes ratios. It also allows estimating the difference in taxes paid using alternative taxation schemes and the impact of the availability to the citizens of such schemes in tax revenues.

Technical information

The model was built using Python with learning libraries in a specialised multiprocessor system for resource-intensive calculations. It is integrated with the existing data pipelines for allowing a broader sharing of the information and integration of the data in different types of statistics and studies. Various data sources were used to develop the system. Mainly, the model was trained using the information provided in the taxation forms containing real income data. In particular, the data from taxpayers who first paid their taxes by module, but then decided to change their paying scheme to income-based and vice versa were extremely relevant.

Drivers and challenges

- A training programme for the staff has been organised in technological foundations
- The planned collaboration with the Spanish Statistical Office will be extremely important for adjusting and validating the results of the model
- Management strongly promoted the exploration of new tools and techniques and allowed staff to be trained in new techniques

Estimation of income for those paying by modules

Spanish Tax Agency

Risks and mitigation measures

- Taxpayers' data protection regulations and some internal organisational regulations on statistical secrets needed to be taken into account. For instance, all data is subject to an anonymisation process before they are used in statistics and other data products
- The information from the AI system was not linked to any decisions but only used as complementary information to assist decision making
- Special measures were taken to mitigate any potential ethical risks. These include taking special care in the selection of data, to achieve fairness of the model.
- The adopted techniques were selected with a specific focus on the explainability of the system

Curiosity box

The Statistical Unit of the Tax Agency, before starting the project, did an evaluation of all the other possible alternatives, to determine if AI was the best option. For example, they explored the possibility of using traditional statistical methods, starting from the original formulas developed for the calculation of the profit and linking them to different more objective measures. Moreover, they also considered to perform a survey to obtain actual measures of income and estimate their current correlation with "modules". In the end, AI was selected as the option with the best cost-benefit ratio. Despite this, the idea of using survey data was not completely abandoned: complementary to the AI system a small survey is planned to validate the outcomes of the model. This testifies the importance of not fully relying on AI but considering it as one possible option that needs to be complemented with other approaches.

Useful links

- https://sede.agenciatributaria.gob.es/Sede/estadisticas.html

Contact details

Félix Jesús Aragonés Arribas
felixjesus.aragones@correo.aeat.es

S.G. de Estadísticas - AEAT
estadisticas@correo.aeat.es

GETTING IN TOUCH WITH THE EU

In person

All over the European Union there are hundreds of Europe Direct information centres. You can find the address of the centre nearest you at: https://europa.eu/european-union/contact_en

On the phone or by email

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

- by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),

- at the following standard number: +32 22999696, or

- by electronic mail via: https://europa.eu/european-union/contact_en

FINDING INFORMATION ABOUT THE EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website at: https://europa.eu/european-union/index_en

EU publications

You can download or order free and priced EU publications from EU Bookshop at: https://publications.europa.eu/en/publications. Multiple copies of free publications may be obtained by contacting Europe Direct or your local information centre (see https://europa.eu/european-union/contact_en).

KJ-NC-31088-EN-N

The European Commission's science and knowledge service

Joint Research Centre

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.

EU Science Hub

ec.europa.eu/jrc

@EU_ScienceHub

EU Science Hub - Joint Research Centre

EU Science, Research and Innovation

EU Science Hub

doi:10.2760/481674

ISBN 978-92-76-53435-8

Publications Office of the European Union