DIGITAL AND POPULATION DATA SERVICES AGENCY

AI Management and Tips

VAHTI Best Practices Support Material

13 June 2024 version 1.5

VAHTI-secretariat

13.6.2024

Table of Contents

AI Management and Tips

1. Introduction

This support material has been prepared for public administration organisations to promote secure work and activities. The VAHTI Good Practices support materials are based on recommendations and good practices compiled by the expert groups of the Finnish Public Sector Digital Security Management Board (VAHTI, https://dvv.fi/vahti) in the areas of risk management, continuity and preparedness, competence development, ICT services, information security and data protection. By complying with these, we also promote the implementation of cyber security.

The VAHTI Good Practices support materials are primarily intended for public administration organisations, but they are freely available to any organisation. We hope that if you develop or improve these materials, you will also give feedback to help us further develop the content. We welcome suggestions for improvements and corrections, and we will release an updated version once we have received a sufficient amount. You can send feedback: digiturva@dvv.fi – enter ”AI Management” to the subject line.

The first version of this document was formed by using artificial intelligence among other things, and it was published on 12 September 2023. This version 1.50 contains updates, that concern in particular the features and other specifications provided by the latest AI services.

This version, published on 13 June 2024, is based on the updated version on the basis of the preparatory working group's preliminary survey open to all.

In this material, we focus primarily on presenting the opportunities and threats of AI services that utilise large language models (LLMs), as they have now become more substantially common over a period of just over 1.5 years and their pace of development has been unprecedented in terms of new features and opportunities.

We utilise artificial intelligence tools in the production of the VAHTI good practices support materials as we have with previous versions of this document. All content will go through a review and comment process.

Dissemination of publication: Digital Security Publications | Digital and population data services agency (dvv.fi)

2. Why has artificial intelligence emerged so quickly and strongly?

Artificial intelligence has been under discussion for nearly 70 years.

”The term artificial intelligence was first coined by John McCarthy in 1956 when he held the first academic conference on the subject. But the journey to understand if machines can truly think began much before that. In Vannevar Bush’s seminal work As We May Think he proposed a system which amplifies people’s own knowledge and understanding. Five years later Alan Turing wrote a paper on the notion of machines being able to simulate human beings and the ability to do intelligent things, such as play Chess.”[1]

and this translated by the AI service without human intervention as an example

<translated to Finnish and then back to English>

AI services are by no means new, OpenAI published a press release on its website on 5 March 2021 on the Generative Pre-trained Transformer (GPT) model they had developed.

“Nine months since the launch of our first commercial product, the OpenAI API, more than 300 applications are now using GPT-3, and tens of thousands of developers around the globe are building on our platform. We currently generate an average of 4.5 billion words per day, and continue to scale production traffic.”[2]

On 30 November 2022, OpenAI published the ChatGPT 3.5 service that spread at a record speed at the end of 2022 and launched a new era in the utilisation of AI.

Version 4.0 of the service was published on 14 March 2023. After the release of the version, the service has been updated with the addition of a plugin functionality, which enables the utilisation of different interfaces.

On 13 May 2024, OpenAI released version 4o, for which significant new features have been promised in autumn 2024, including advanced voice mode, support in 50 languages, a more advanced multimodal user interface that will also enable the interactive unloading of video content and advanced interpretation functionality. Most of these features are also promised to the users of the free-of-charge service. [3]

The example above has been picked out of an AI service developed by OpenAI, but at the same time we must keep in mind that the same kind of development is underway in all major companies developing AI services, including

Google Gemini	https://gemini.google.com
Meta	https://www.meta.ai/

[1] The History of Artificial Intelligence (washington.edu)
[2] GPT-3 powers the next generation of apps (openai.com)
[3] Introducing GPT-4o and more tools to ChatGPT free users | OpenAI

Microsoft CoPilot https://copilot.microsoft.com
free-of-charge versions that e.g. link to cloud services
Claude Sonnet https://www.anthropic.com/news/claude-3-5-sonnet

In addition to these, hundreds, even thousands of companies have added services that can be classified as AI services to their service range, including chatbot-type assistants or interfaces with large language models.

During the past year, significant qualitative improvements have been achieved, especially in the production of images, and the same leap is expected next for videos produced with artificial intelligence.

It is also expected that the use of natural language as a user interface for AI services will increase, which may have a significant impact on, for example, the provision of services for all of us.

2.1 Artificial intelligence is one of the most significant technological developments

IBM released the PC in August 1981. Data networks that link computers to one another and, in particular, the Internet and web browsing in the 1990s were significant developments in IT and ICT. In the 21st century, the spread of mobile technology and smart devices has shaped our way of using digital services independent of time and place.

It is not yet possible to fully understand and evaluate the significance of artificial intelligence. One of the Osmo W. Wiio's laws concerns predicting the future: "The near future is overestimated, and the distant future is underestimated." We may finally be at the point when, after having overestimated the potential of artificial intelligence for the last few decades, we will soon underestimate its importance to the near future, no matter how we try to understand it.

Key facilitators of artificial intelligence include the development of the performance of ICT technology (computing power, speed of data connections, capacity of storage systems) and a significant decrease in the prices of technological devices. AI services would not be possible without the **Moore's law** being realised[4]

Moore's law, named after Gordon Moore, one of the founders of Intel, which manufactures core microprocessors, is the observation that the number of transistors in an integrated circuit doubles about every two years, which, when roughly generalised, also doubles the performance of computers. It has been a fairly accurate forecast for decades, but there is more and more discussion in expert circles about when and how this trend will come to an end. Many estimates now assume that Moore's law

[4] What Is Moore's Law and Is It Still Relevant in 2023? (makeuseof.com)

could cease to be true by the end of the 2020s, but there is a lot of uncertainty about this.

Some technologies, such as **quantum computers** or **new semiconductor materials**, will be explored as possible ways to continue to increase performance after this. It is important to note that even if the density of transistors were to no longer increase, IT can continue to progress in other ways. For example, **software algorithms**, **AI,** and **specialised circuits** (such as graphics processing units or AI chips) improve performance or computing efficiency even if hardware performance no longer doubles every two years.

An excellent example of this is the US company Nvidia[5], which has become one of the fastest growing technology companies in the world over just the last couple of years. The company has managed in particular to create graphics card processors, originally used in gaming and demanding graphics processing, to speed up AI computation.

In addition, the development of artificial intelligence has been promoted by a significant increase in the amount of data and the development of algorithms used by artificial intelligence.

One of the challenges related to the increase amount of data is related to the increase in the amount of information created by artificial intelligence. If artificial intelligence does not actually create anything new, is there a risk that the information available to us will be "dumbed down" or that we will become dumber? NOT necessarily because artificial intelligence provides new perspectives and thus, we will be able to act by utilising our key resource as living people, our creativity, to produce new innovations, ideas and theories.

Even if the AI boom launched at the end of 2022 were not to continue at the same intensity in the coming years, it has nevertheless triggered broad-scoped change. If the development of new and existing services were to slow down, it will not halt the change that has already started; but as we write this, that does not seem to be something in sight in the near future.

[5] Nvidia - Wikipedia

3. Tips for personnel on how to test and utilise AI services

The instructions below are intended to be taken into account in the utilisation of open, publicly available, generic AI services, especially those based on large language models (LLMs). If an organisation uses a service purchased separately, it must provide instructions on how to use the service. It is also advisable to review the issues raised here.

3.1 Five key points in the use of artificial intelligence

Before rushing off to utilise artificial intelligence or, at the latest, after your initial enthusiasm has worn off, note the following four things:

1. Understand what artificial intelligence is and how it works
2. Maintain your competence as a user of services
3. Identify what data can be entered into the service
4. Check the data produced by the services before publishing or utilising it
5. Prepare for abuses by cybercriminals and other hostile actors

3.1.1 Understand what artificial intelligence is and how it works

Understanding how AI works makes it easier to assess the factors, both the risks and the opportunities associated with the use of AI services. In this support material, we focus strongly on AI services implemented using large language models (LLMs), as they have now brought the utilisation of AI to a completely new level.

OpenAI ChatGPT-4 itself defines "Artificial Intelligence (AI) is a field of computer science that focuses on the development of smart machines, software and systems. The aim of AI is to create systems that are capable of performing tasks that normally require human intelligence. These tasks may include image and speech recognition, learning, planning, problem solving and decision-making."

As the AI architecture and operating model may differ significantly depending on the service and design used, the organisation must fully understand and identify the operating model in question. Understanding one operating model and service does not guarantee that the same lessons apply to another service.

An AI service requires the following to work:

Data
Artificial intelligence and machine learning are largely based on data. This can include historical, real-time, publicly available, or a company's in-house data. It may also be data in an online service protected by access rights, for example in a social media service. The data type can be structured (e.g. in tabular format) or unstructured (e.g. images, text).

Algorithms (operating logic)
Algorithms are the heart of machine learning and artificial intelligence. They learn models and relationships from data and make forecasts or decisions based on them.

Different algorithms exist for different purposes. In an ordinary ICT service, the algorithm always works in the same way, for example in table calculations 1+1 = 2 and 4*5=20. On the other hand, while the answer given by AI services should be correct, it is possible that when asked a hundred outcomes, one of them will be 3 or 25. Indeed, the usability of AI services is significantly limited by the fact that the information they produce must always be reviewed.

It is also worth remembering that the responses are influenced by a previous discussion. In the future, it is expected that more sophisticated services will have enough memory available to remember all conversations.

Machine learning models
Algorithms use data to train machine learning models. Models are those that make actual forecasts or decisions when they receive a new input.

Teaching data
AI teaching data consists of the information used to train the AI model. This data serves as teaching material to help the model identify images, make predictions or perform other tasks. The quality (correctness) and quantity of teaching data are critical factors for the functioning of the model. Teaching data is only one part of the AI training process, but a very critical one.

Computation resources
AI services often require significant computing resources, especially for processing large amounts of data and training machine learning models. This can happen locally or in cloud-based services.

Interface
Users of an AI service need methods to communicate with the service. This can happen, for example, through a web interface, a mobile application or an API, and in the future, more and more often by voice, images or video.

Architecture and ICT infrastructure needed to produce services
AI services need a well-designed architecture and ICT infrastructure that enables efficient data processing, computing and service use. They can include databases, servers, network connections and other technologies.

Information security and data protection, **privacy protection**
AI services must take information security and data protection into account, in particular when processing personal data. For example, encryption, access rights and possible pseudonymisation or anonymisation of data subject to data protection enable the implementation of information security and data protection. It is particularly important to know whether or not the data processed in the AI service ends up as part of the service's teaching data. This will have a significant impact on whether, for example, confidential or personal data can be processed in the service. An organisation may procure an AI service that also enables the safe processing of the aforementioned data.

Ethics of AI
The ethics of AI refers to principles and guidelines that guide the development and use of artificial intelligence so that they are ethically acceptable and promote the well-being of society. The ethics of AI aim to ensure that AI technologies are developed and used in a way that is fair, safe, transparent and responsible.

- *Fairness and equality*
    - Ensure that AI systems do not discriminate against users on the basis of such things as race, gender, age or other personal characteristics.
- *Transparency*
    - AI must work in a way that can be described and understood so that users and stakeholders can understand how and why AI makes certain decisions.
- *Accountability*
    - AI developers and users must be accountable for the use of AI and its impacts. This also means that mechanisms are in place to take accountability if artificial intelligence causes harm.
- *Security and reliability*
    - AI systems must be secure to use, and they must function reliably in different conditions. This also includes risk management and assessment.
- *Privacy and data protection*
    - Artificial intelligence must respect the privacy of individuals and ensure that personal data is processed appropriately and securely.
- *Doing good and avoiding harm*
    - AI systems should promote the good of society and minimise possible adverse effects. This means developing and using artificial intelligence in ways that benefit as many people as possible.

Also take the following into account when using AI services:

Training AI

Language models have been taught, i.e. ‘trained', by entering large amounts of text containing a wide range of books, articles, and websites. During the training, AI learns to statistically identify how words and phrases are interlinked and how they form meaningful messages. In addition to this, training may also include additional training carried out as human work, in which answers that the party who owns the artificial intelligence does not, for one reason or another, want the model to produce are eliminated. It is therefore important for the user to understand that AI does not

produce answers directly based on the data, but that the outcome is influenced by such factors as the values of the company providing the service, its business logic and applicable legislation. Thus, the result is not necessarily neutral. The same should also be considered when using search engines, as search results can be influenced by various means.

Generating a text or other response

When artificial intelligence receives a prompt, such as a question or phrase, it tries, based on what it has learned, to select the next word that is statistically best suited to the textual context. AI repeats this process again and again until it has produced the entire answer or paragraph.

Artificial intelligence has no understanding or awareness

Although the AI service can produce text that looks understandable or conscious, it is important to understand that AI does not really have an understanding or awareness. The information it produces may look correct and credible in a logical sense, but its content may be complete nonsense. For this reason, it is very important that the information produced by AI is reviewed and verified by an expert with sufficient expertise. It does not "know" or "understand" information in the same way as a person. It just creates models based on what it has seen before and is mathematically closest to the correct answer.

An AI service operates in a simplified manner, similar to predictive text input used in smart devices which aims to predict the word you type on the basis of the characters you enter. Instead of a single word, AI predicts significantly larger entities based on the material it was trained with.

Limitations

There are limitations on the use of AI services. It will not always produce fully accurate or reliable information, and the answers may vary even if the question is the same. It also does not necessarily remember previous discussions, nor does it have the ability to understand or assess a person's emotions.

3.1.2 Maintain your competence as a user of services

Today, we rarely have to learn completely new types of digital services or tools, as some of us have used Windows, the Internet, email and office programs for perhaps even decades. Their new versions rarely bring significant changes.

On the other hand, AI services will develop at an unprecedented rate, which is why everyone should participate in related training and actively familiarise themselves with the new or developing opportunities offered by the services. If anything, it is the utilisation of artificial intelligence that encourages someone to be digitally courageous; not recklessly but by identifying threats and controlling risks!

3.1.3 Recognise what data can be entered into the service

Never enter sensitive, personal or confidential data into AI services unless the service has been classified for this type of use by your organisation. It is also advisable to avoid entering other identifiable data related to the organisation, if it is not known how the service will process it. In this case, there is a risk that such data may end up, for example, as part of the service's teaching data and reveal information on the organisation's activities to outsiders.

Recognise the threat that if organisation-specific data ends up in the service's teaching data, such as data that includes a name or otherwise identifiable, someone else can access this data by directly asking "What kind of strategy does xyz company have?"

Make sure that you know how to identify and distinguish the public data used in the services from prohibited data. You can utilise the following model in the secure processing of data: Identify Data, Facilities, and Tools – link to the material in Finnish here. In this model, AI services can be seen as a tool among other digital services.

3.1.4 Review the data produced by the services before publishing or utilising it

Remember to check the results produced by AI services. Although AI can automate many processes, its results must always be reviewed before they are published or otherwise utilised.

Errors in the services may also be reflected as ethical problems in the data they produce. Ensure that the services or the data they provide do not promote discrimination or injustice.

Please note that AI cannot be used in public decision-making without a separate approval process, as automatic decision-making requires e.g. that the models and algorithms used are public. This is often not possible for AI algorithms because the model is modified according to the training material and is not always transparently available. Legislation on this is currently being planned.

Also save the prompts you have used to produce material (text, images, videos). You may later need to check how you produced this information using the AI service.

As the operations of AI services are often closed, it is usually not possible to check the algorithms behind them. The accuracy of information produced with this type of "black box" must be checked particularly carefully.

3.1.4.1 What can cause hallucinations?

AI services may produce "hallucinations", in which data produced by the service appears to be correct in principle, but in practice it may be completely incorrect. The result may reference sources that appear to be correct, but which cannot be found or

contain errors.

Lack of context

AI models do not always fully understand the wider context or the background of certain questions. This may lead to them providing information that sounds right but is not accurate or correct.

Limited availability of information

Artificial intelligence models are trained with huge amounts of data, but they cannot know everything. If the model does not have sufficient relevant information on a particular topic, it may fill the gaps by inventing the information as if "on its own".

Statistical reasoning

AI models are based on statistical reasoning. They predict the next word or phrase based on previous models, but they do not know whether or not the prediction is correct.

Ambiguity

Natural language is ambiguous, and many words and phrases can mean different things in different contexts. This may lead to the model producing answers that are technically correct, but in the completely wrong context.

Teaching data / Training data

AI models are trained with huge datasets that contain both correct and incorrect or inaccurate information. The model is not always able to distinguish between these, which may lead to the generation of incorrect data.

Generative models

GPT models are generative, which means that they create answers on the fly by predicting what a sensible answer could be. However, this process does not guarantee that the information created is true.

Incomplete instructions

If the user's question is unclear or ambiguous, the model may "guess" the answer based on previous similar questions. This may result in incorrect information.

3.1.5 Prepare for attacks and abuses by cybercriminals and other hostile actors

Using artificial intelligence in services and equipment offers us new opportunities to develop services and equipment that are more interactive and user-friendly.

However, it should be understood that alongside the opportunities that it offers for the benefit of humanity, it is just as easy or sometimes easier for cybercriminals and other service abusers to exploit these opportunities.

All the positive opportunities that we can implement through AI services can therefore be turned against us. It should be noted that cybercriminals are not subject to any laws or ethical obligations, and it is therefore expected that over the next few years we will experience an increasing number of scam campaigns, other attacks and violations that will make extensive use of artificial intelligence. Below are some examples of the areas that, when implemented with AI services, offer more advanced or novel opportunities for abuse:

3.1.5.1 Unlimited opportunity to produce and analyse text in different languages at a high standard

Implementing digital scam campaigns in which:

- The content of the message is as authentic as possible, or the content of the messages is modified to make it difficult to identify.

    - This can also be targeted more precisely than before in more specified attacks with the help of data collected on the victim.

- The ability to combine the content of different texts to analyse and thus identify different errors or vulnerabilities.

    - This also provides an opportunity to identify something that is critical for the organisation or individual or even a confidential data set by combining public data.

As generative artificial intelligence enables natural discussion, this provides an opportunity for cybercriminals to create, for example, a seemingly human-like digital scam bot that searches for victims in online services, such as social media platforms or by phone. This also provides a very positive opportunity to create various support services, for example for the elderly, but at the same time creates a frightening opportunity for abuse.

AI services have significantly contributed to improving the quality and general development of text and voice-based language translation services. As a result, it is expected that the text used by cybercriminals and other scammers will continue to improve and that it will no longer be possible to identify it due to grammatical or spelling errors.

3.1.5.2 Possibility to produce and evaluate programme code

An attacker can use this to generate their own code or to find vulnerabilities in an item.

- For example, malware can adapt and change its behaviour polymorphically to remain undetected for as long as possible.

- Cybercriminals can build a "bot" that automatically journeys through the Internet searching for and keeping notes on the vulnerabilities they know and after finding such a vulnerability launches an automated attack intended for the vulnerability in question.

3.1.5.3. Implementation of deep fakes

Cybercriminals can already capture material from existing videos, audio files and photos quite easily to produce genuine-looking photos, voices and videos.

If we now warn users that a cybercriminal impersonating their family member or friend, for example, may approach them with an instant message requesting information, this message may in the future take the form of an audio file or even real-time speech.

The same applies to the authenticity of videos, which we have seen on the Internet in authentic-looking videos of known people which have proved to be fake. In early 2024, a cybercrime was reported in Hong Kong, where a company's financial expert was scammed at an online meeting with the use of deepfake videos and characters, and the company lost $25 million in this crime.[6]

This technology makes it easier to influence information and spread disinformation as well as to blackmail people. These technologies will be increasingly utilised, for example, in electoral interference.

3.1.5.4. Recognition and analysis of objects, characters, and faces

AI services can be used more accurately than ever before to identify, such things as emotions from people's faces and other behaviours that can then be used to exploit the people. Such profiling is prohibited under the EU Artificial Intelligence Regulation, but unfortunately the Regulation only applies to the EU area or operators providing services in this area. As previously stated, cybercriminals do not observe laws or other ethical principles.

If you are now using an application to identify e.g. animals, plants, or music tracks, the same functionality can be used to identify anything. For example, modal input to AI services makes it now possible to take a photo, in the future to record a video, and ask the service to analyse the things, objects and events shown in the image.

[6] https://edition.cnn.com/2024/02/04/asia/deepfake-cfo-scam-hong-kong-intl-hnk/index.html

3.1.5.5. 24/7 capability of AI services

Like all technologies, AI services can be harnessed tirelessly to carry out the task assigned to them.

For example, a cybercriminal may launch an Internet-based programme that collects and identifies any confidential data found in the services or technical vulnerabilities related to the services and, for instance, initiates an automatic blackmail process targeted at the organisation in question.

Such services have been in use for a long time, but now AI brings better opportunities for abusers to identify and exploit them and automate all of this through AI.

Similarly, the attacker will be able to use this capability to implement more advanced and targeted denial of service attacks than at present.

The five examples above are some of the possible options for how cybercriminals and other abusers can use artificial intelligence. Most of the methods have been in use for a long time, but we will also encounter new forms of cybercrime and significant qualitative development of existing methods.

3.2 Also check these points if you intend to use AI services in your free time

These recommendations can also be applied to the procurement of AI services in organisations.

3.2.1 How familiar and well-known is the service provider?

An enormous number of new partly start-up-type companies have entered the market. Their services may suffer from urgent implementation and their lifetime may be short. It is a good idea to search for more information and experiences regarding the service that are as reliable as possible before using it more extensively or to purchase a paid right of use. Several services initially offer a fixed-term free trial period, which is definitely worth taking advantage of before purchasing a service. We recommend using the services of the most well-known service providers, as they have the resources to develop and add new features to them and to see and correct any errors and vulnerabilities faster than smaller operators.

3.2.2 Watch out for fraudulent services and check terms and conditions of use

Cybercriminals have also shown a huge interest in AI services. Do not search for services using "Google" or download services directly from the links offered to you. Download services from their manufacturers' websites and on smart devices from Google Play app store or Apple AppStore. These stores may also have applications that do not function as they claim. For this reason, it is advisable to read e.g. reviews for the services.

Also, make sure that there are no unexpected costs or longer-term payment commitments associated with the purchase or use of the service before you are sure you are willing to pay for the longer-term use of the service in the form of a monthly fee.

3.2.3 Read the service settings and its user agreements through carefully

Check what kind of rights the service provider gets for the (teaching) data to be entered in the service and how it will utilise the data accumulated in the service - including on its use and users.

Make sure whether it is possible to limit the teaching data to your own use, so that it does not spread to other service users. Or if this will require e.g. a separate, paid for version.

Verify how you can delete data added to the service after use. Please note that any questions you ask the service (prompts) may be made available to other people if, e.g., you demonstrate the use of the service to others on your device.

Read and learn how to ensure in services intended for the production of images, videos and music that these outputs are not shared openly with everyone (public) unless you explicitly want to do so.

3.2.4 Consider which email or username you will choose to use in the utilisation of the services

Consider creating a new email address for testing AI services. Many smart devices, such as Apple's iOS, offer the "Hide my email" feature[7]. In this case, when creating a username, you will be provided a unique email for the service, which directs the emails sent by the service to your personal email account. This will give you anonymity in the service in question.

Similarly, you can add a + (plus) character to the end of your Gmail address and follow this by any characters you wish. For example, aku.ankka@gmail.com will also receive email from the addresses aku.ankka+artificialintelligence@gmail.com or aku.ankka+pretty.cool@gmail.com. Similarly, you can try a dot because the Gmail service doesn't recognise any dots in emails, as messages sent to a.ku.ank.ka@gmail.com will also be delivered.

3.2.5 Make sure you have access rights to the data you use to train AI

For using and sharing of information, for example, copyrights and access rights must be taken into account. At your workplace, you may have access to data or reports produced for your organisation the disclosure of which to third parties is prohibited in your employment contract. In this case, you do not have the right to upload the report in question to the AI service provided to you by your organisation if the service utilises all data as its teaching data. Similarly, it should be verified whether such a report

[7] Kätke osoitteeni -osoitteiden luominen ja hallitseminen iPhonen Asetuksissa - Apple-tuki (FI)

can be uploaded even to an AI service that does not export the data into teaching data.

4. Examples for organisations on the creation of an administrative AI framework

In this chapter, we have compiled a few examples of the things and documents that organisations should consider drawing up as part of the process of using their AI services.

We want to emphasise that the organisation's management must now **actively take part in creating and developing** the "AI framework" needed by their organisation, through which the organisation will implement AI services in a controlled manner. Otherwise, there is a risk that the personnel may get frustrated and utilise leisure time equipment, services and artificial intelligence to carry out their work tasks, i.e. the so-called Shadow AI services.

4.1 Personnel competence of key importance

ICT services and digital devices have made it possible for us to utilise digital services and perform tasks in a new way. In the future, the utilisation and rapid spread of AI services will be a challenge to some users, which will be compounded by the rapid development of AI services.

As was when IT became more common in the 1990s and 2000s and digitalisation developed in the 2010s, there will be a risk that some people will not want to learn another new technology or service. Or some users will feel that new types of AI services are not for them.

AI services can help all age groups and it is very likely that we will be able to offer new, easier-to-use and more intuitive AI services and devices equipped with artificial intelligence, for example, to facilitate the everyday lives of older people. One possible megatrend in the late 2020s will be AI services, which will probably be combined with different service robots and avatars operating in metaverse environments in the future.

This will certainly have an impact on how difficult or easy it will be to maintain a separation between the physical and virtual world in the future. The mixing, interaction and understanding of these two different operating environments will become significantly more difficult to understand.

One of the challenges of new services and equipment is the need to learn to let go of old practices and boldly introduce new types of services and equipment.

Organisations must actively train their personnel on the opportunities and changes related to the utilisation of AI services. It is important to talk openly about both positive experiences and the problems and challenges experienced in the development and utilisation of services.

4.2 Things to take into consideration in an AI strategy

An AI strategy is an organisation's plan for how it intends to use AI to achieve its goals and develop its activities. This can include all kinds of issues concerning the use of AI in daily activities, such as customer service or larger initiatives such as product development or the creation of new business models.

Below is a list of topics on which the AI strategy should take a stand:

1. Definition of objectives and priorities

Why and how does the organisation want to utilise artificial intelligence? What are the main objectives and priorities? Objectives should be in line with the organisation's broader strategy and objectives.

2. Ethical principles

A sound strategy defines the organisation's ethical principles and rules for using AI. Previously, the following perspectives were highlighted in this material:

Open and Transparent Use **|** Fairness and Non-Discrimination **|** Data Protection and Privacy **|** Security and Reliability **|** A Human-centred Approach **|** Ethical Planning and Use **|** Continuous Monitoring and Assessment **|** Participation and Collaboration

3. Data management

AI needs high-quality data to function efficiently and reliably. A strategy should define how data is collected, stored, analysed, and shared within and outside the organisation. The strategy must also define what kind of teaching data is used to train AI if the organisation is handling the AI training process and how the quality of results can be measured and monitored. Monitoring related to the quality of teaching data must also be included in this process.

4. Development of competence and resources

What kind of expertise and resources are needed in the organisation for the utilisation of AI? Where and how can these be obtained and how will these competences and resources be developed?

5. Choice of technology and its deployment

What kind of technology does the organisation intend to use and how will it be deployed? The strategy should include a plan for the selection, procurement, testing and deployment of technology. The services to be deployed should support the architectural or other principles already in use in the organisation.

6. Collaboration and partnerships

How does the organisation intend to engage in cooperation with other public organisations, the private sector, universities and others?

7. Measurement and monitoring

How does the organisation intend to measure and monitor the impacts of using AI and its results?

After the strategy has been created, it is important that it is a living document that is regularly updated to reflect new knowledge, experiences and changing circumstances. In addition, the implementation of the strategy must be monitored and supported with sufficient resources.

4.3 Draft on matters to be noted in AI policy

This draft policy is designed to guide an organisation in the use of AI, ensure ethical and responsible practices and promote continuous improvement and learning.

1. Setting strategic guidelines

We define concrete, measurable goals to help us track the progress of our AI projects and assess their success. We always strive to ensure that our efforts support the overall strategic objectives of the organisation.

For example, if our objective is to improve customer service through AI, we set a concrete goal on how much we intend to reduce the processing time of customer service requests with the use of AI over the next year or on how much customer satisfaction needs to improve.

2. Accountability and transparency

We comply with ethical standards in all our AI work. This means, for example, compliance with the principles of non-discrimination, data protection and transparency. We also strive to explain the functioning of our AI models as openly as possible.

We are developing practices that will require all AI models to be tested for discrimination prior to deployment. We will also strive to publish the principles and algorithms for decision-making in our models as far as possible. We ensure that the quality of our teaching data is monitored with the aim of minimising misinformation and hallucinations, possible discrimination or bias related to the results produced by artificial intelligence.

3. Utilisation and protection of data

We collect and use data responsibly, observing all applicable laws and standards. We also strive to ensure the quality of data, as we understand that it is the key to

efficient and reliable use of artificial intelligence.

We store all information we collect securely and comply with data protection laws. In addition, we only use data to train AI when we have the appropriate rights and permissions to do this.

4. Personnel training and competence development

We invest in the artificial intelligence expertise of our personnel. This means that we organise training programmes, mentor or hire experts, etc.

As an example, we provide annual training on AI for all our employees and organise in-depth workshops for those working in AI projects.

5. Selecting the most appropriate technology

We select AI technologies with care, considering their security, cost-effectiveness, usability and suitability for our organisation. We are open to different technologies and choose the best tool for each task.

At the beginning of each AI project, we assess which AI platform or tool we have in use is best for the task in question. We take costs, support resources, compatibility with other systems, and user-friendliness all into consideration.

6. Building a collaboration network

We actively seek partners from both the public and private sectors, including educational institutions and research institutes. We believe that together we can achieve more and create better AI solutions.

We work together with stakeholders to develop and share good AI practices.

7. Continuous evaluation and improvement

We regularly evaluate the performance of our AI projects and make the necessary corrections. We use both quantitative and qualitative indicators and learn from both successes and failures. We aim to act iteratively, continuously improving, and optimising AI applications and processes.

We use tools such as customer satisfaction surveys and business data to assess how well our AI models work and what impacts they have had. We also hold regular feedback meetings with AI project teams so that we can learn directly from those who work with AI on a daily basis.

4.4 Implementation of risk management

Risk management must be implemented for artificial intelligence just like it is for other services. The recommended method is to use the operating model used by the organisation, which is adapted to the AI services used. This is also one prerequisite from the perspective of the EU AI Regulation.

The organisation should identify such things as high-risk and low-risk targets related to the use of AI services:

- Exercise of public authority,
- Utilisation of personal data,
- Use of confidential information,
- Issues related to the information security of the service, in particular data protection,
- Copyright-related issues,
- Risks related to the accuracy of information

Artificial intelligence applications may produce information that is completely wrong, so risk management must take into account how incorrect information will be corrected and how the information produced by artificial intelligence will be reviewed before it is published.

Some of the risks are likely to be such that accepting them as such or even with additional controls may be challenging for the organisation. For this reason, it is worth focusing on tasks where the risk is low or can be managed, especially in the early stages. These include:

- processing of public data,
- communication and provision of information,
- production of general, public materials.

One of the key risks is linked to the use of a service provided by an external service provider with which it is usually not possible to conclude any agreement. The terms and conditions must be accepted as such, and it is not possible to assess the security of the services.

One of the key opportunities for managing risks is to use an AI service built for the use of public administration, such as one developed and located in Finland.

Risk categories related to artificial intelligence

Data-related risks

The risks may relate to data quality and accuracy, the protection of data privacy and security, and the ethical collection and use of data. Inaccurate or incomplete data may lead to erroneous results and its misuse to privacy violations or other problems.

Algorithm-related risks

Risks may be related to how artificial intelligence is used, but they are also caused by algorithm delusions, over-optimisation or over-adaptation, as well as a "black box" problem in which it is difficult to understand or explain how the algorithm works.

Use-related risks

These risks are linked to such things as to the misuse of an AI system, such as its use for harmful purposes or its use without sufficient understanding of how it works. Accepting incorrect AI responses for further use is also a risk associated with use. This risk can be reduced by having an independent party verify the answers, if possible.

Risks related to liability and legislation

Who is accountable if artificial intelligence makes a mistake or causes harm? This is a complex issue with many legal and ethical factors. In practice, the organisation's management has overall accountability for these services either produced by the organisation itself or purchased for its use.

IPR risks (Intellectual Property Rights)

The use of generative AI involves several IPR risks that organisations should consider. For example, if an AI solution is trained with data containing copyrighted material, a copyright infringement may occur. It is important to ensure that training data is correctly licensed.

It is clear that artificial intelligence will change and stimulate discussion on IPR rights in the future. New IPR systems may be needed to address the unique challenges brought about by AI.

Cybersecurity risks

From the perspective of ICT services, AI systems can be vulnerable to attacks, such as data-based attacks, denial of service attacks or malware. These risks can be mitigated by ensuring the protection of the AI system and regular security audits. In this respect, the technical, administrative, and physical security of AI systems must be ensured in the same way as all other ICT services.

Social and ethical risks

Artificial intelligence has the potential to change society in many ways, and not all of

these changes are necessarily positive. When implementing AI, it is important to take into account its potential impact on people and society, including issues of discrimination, fairness and human rights.

Risks related to teaching data being revealed

If an organisation teaches AI itself using its own training data, in some cases the AI model can be used to access the original training data. If the data contains the organisation's confidential data, there is a risk that these will be revealed to external parties.

Risk management methods

All the above-mentioned risk categories must be reviewed as part of the risk assessment and adequate controls must be described and implemented at the level required to manage the residual risk. If the risk cannot be reduced to an acceptable level, the service cannot be implemented or deployed.

Legislative risk management

The organisation must ensure that its AI solutions observe all applicable laws and regulations.

Ethical risk management

The organisation must take ethical aspects into consideration and ensure that its AI solutions are ethically sustainable.

Managing information security risks

Information security strategies and policies must be in place for the management of risks.

4.4.1 EU Artificial Intelligence Regulation

Objectives and application

‘AI-system’ means a machine system designed to operate to varying degrees of autonomy, which may adapt after deployment and which, on the basis of input data received, may produce outputs, such as forecasts, content, recommendations or decisions that may affect real or virtual environments, in order to achieve specific explicit and implicit objectives.

Risk-based approach

The regulation defines a four-step scale, the most important of which is to identify what contains ‘too high risk', i.e. the use of these is prohibited.

- **Too high risk, prohibited practices**

This first category includes AI systems the use of which involves risks that are not acceptable. Systems in this risk category are considered to pose such a clear threat to people's safety, livelihoods, and rights that their use is totally prohibited.

- Subliminal techniques that people do not consciously detect or techniques that are intentionally manipulative or misleading. The aim is to distort or substantially distort the behaviour and decision-making of a person or a group of people.
- Thus, taking advantage of the vulnerabilities of a particular person or a group of people (age, disability, special situation).
- Social scoring, i.e. a system that assesses or categorises people or groups and leads to adverse treatment or unreasonable consequences.
- The prediction of crimes with risk assessments of a person's personality (cf. the Minority Report movie) (with the exception of assessments that act as support function for human assessments).
- Facial recognition by sweeping facial images from the internet or surveillance camera images in an unspecified manner.
- Emotion identification of employees or students (except for medical and safety purposes).
- Biometric identification based on specific information (race, religion, political opinion, etc.).
- Real-time remote biometric identification in public spaces for law enforcement purposes. (Excluding the identification of human trafficking, prevention of terrorist attacks and identification of criminal suspects.)

Other risk groups include:

- **High risk, high requirements**
- **Limited risks of certain services, transparency obligations**
- **Low risk, no requirements, but recommendations**

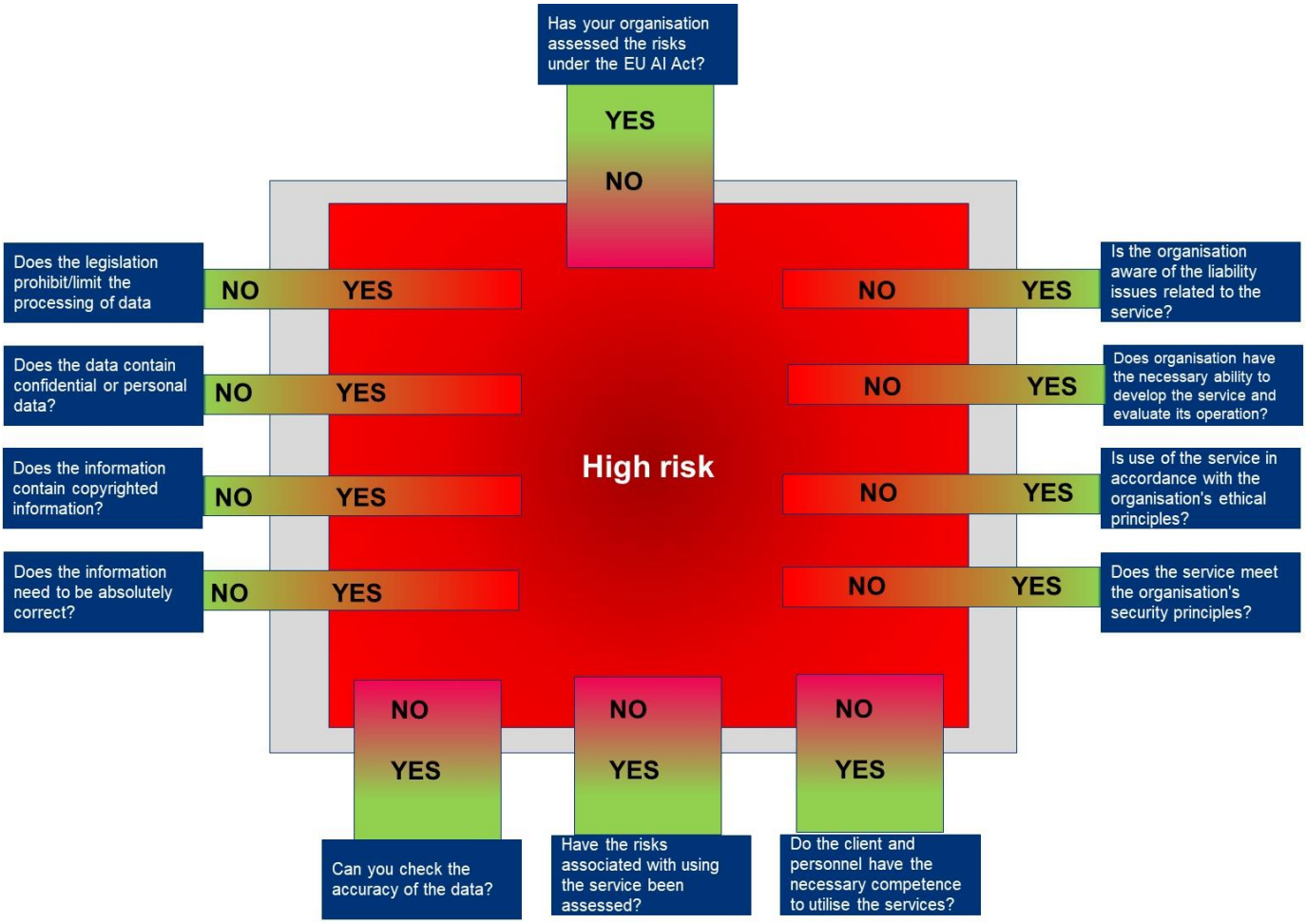
The aim is not to describe this entity in its entirety in this material, but to give an idea of the risk categories of the AI Regulation. For more information, see e.g. Artificial Intelligence Act | Shaping Europe's digital future (europa.eu) and other future support materials related to the Regulation. In particular, it is advisable to look at the application schedule, which includes steps 6 – 9 – 12 – 24 and 36 and sanctions, such as warnings and administrative measures, of which administrative fines can be very significant for companies (EUR 15 or 35 million or up to 3-7% of the total annual global turnover of the company for the previous financial year, whichever is higher in the case).

4.4.1 Example - How can you easily assess the risks associated with the use of artificial intelligence?

The use of an AI service, just like the use of all services produced or implemented by the organisation, requires continuous risk management.

Review the presented questions and provide answers. The more answers are placed in the red (higher risk) area, the more it must be ensured that the deployment or utilisation of the service is possible, legal, and safe, and in accordance with the organisation's policies.

This same risk pool model can be applied effectively in the design and assessment of non-AI services.



5. Legislation and artificial intelligence

Take into account the general laws related to the use of AI and identify specific legislation that applies to your organisation, which may impact the use of AI. Below are a few examples, however, this list is not complete.

5.1.1 Act on the Openness of Government Activities

Laki viranomaisten toiminnan julkisuudesta 621/1999 - Ajantasainen lainsäädäntö - FINLEX ®

22 § Document secrecy

"A secret official document, a copy or a printout thereof shall not be shown or given to a third party or made available to a third party by means of a technical interface or otherwise."

5.1.2 European General Data Protection Regulation (GDPR) and national legislation

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (europa.eu)

i. Lawfulness of processing (Article 6)
ii. Transparent information, communication, and modalities for the exercise of the rights of the data subject (Article 12)
iii. Right to rectification and erasure (Article 16 and 17)
iv. Right to restriction of processing (Article 18)
v. Data protection by design and by default (Article 25)
vi. Data protection impact assessment (Article 35)
vii. Prior consultation (Article 36)

5.1.3 Act on Information Management in Public Administration

Laki julkisen hallinnon tiedonhallinnasta 906/2019 - Säädökset alkuperäisinä - FINLEX ®

15 § Ensuring dataset security

An authority shall ensure, with the necessary data security measures, that:

1) the unaltered state of its datasets has been sufficiently ensured;
2) its datasets have been protected against technical and physical damage;

3) the authenticity, timeliness and accuracy of its datasets have been ensured;

4) the availability and usability of its datasets have been ensured;

5) the availability of its datasets is restricted only if access to the information or processing rights have been separately restricted in the law;

6) its datasets can be archived, as required.

5.1.4 Administrative Procedure Act - Chapter 2 Foundations of good administration

Hallintolaki 434/2003 - Ajantasainen lainsäädäntö - FINLEX ®

- principle of equality (Administrative Procedure Act, section 6)
- principle of intended purpose (Administrative Procedure Act, section 6)
- principle of objectivity (Administrative Procedure Act, section 6)
- principle of proportionality (Administrative Procedure Act, section 6)
- principle of legitimate expectations (Administrative Procedure Act, section 6)

5.1.5 Proposal for a Regulation of the European Parliament and of the Council on harmonised rules for artificial intelligence (AI Act) amending certain Union acts

Newest version approved on 21 May 2024:
https://data.consilium.europa.eu/doc/document/PE-24-2024-INIT/en/pdf

Old version from 21 April 2021

https://eur-lex.europa.eu/legal-content/FI/TXT/?uri=celex%3A52021PC0206

5.1.6 Requirements for the security of AI services (cyber) (information)

Note that the majority of AI services utilise global cloud services. This means that the secure and reliable use of these services must be ensured. An organisation must ensure that the cloud services it uses meet the legal requirements of the organisation concerning the utilisation and procurement of ICT services. In addition to the previously mentioned Act on Information Management in Public Administration, the organisation must identify the special legislation applicable to it and the Cybersecurity Act, which will enter into force on 18 October 2024 (HE 57/2014).

Cybersecurity Act
https://www.eduskunta.fi/FI/vaski/HallituksenEsitys/Sivut/HE_57+2024.aspx

Appendix 1 - Examples of the application of artificial intelligence

AI services are currently evolving very rapidly. Do not judge or underestimate their performance and significance due to individual errors or inaccuracies. This section primarily first and foremost describes new types of potential uses that have developed rapidly in 2023–2024. The use of AI services will become increasingly interesting, as AI services will be able to create new content efficiently and help us in everyday tasks. The following sections describe possible uses.

1.1 The services can process an enormous amount of text, summarising or abbreviating these, and answer content-related questions

Would it not be tempting to ask an AI service "to produce minutes from the recording of the last meeting" or to ask, "what were the decisions made at our last meeting?" or "who spoke about the OpenAI service for the first time at our meeting at the end of last year?"

The organisation must understand what utilisation of such function means, for example, regarding the confidentiality, integrity and accessibility of data and the processing of personal data. In addition, the utilisation of AI services involves completely new requirements related to the correctness, transparency and ethicalness of the information or decisions produced.

1.2 The services can translate text, speech or read text in images and translate it from one language to another

For a long time, there has been numerous advanced translation services that translate from one language to another, and now there are also machine learning or language model-based translation services, the quality of which is constantly improving. As with all other services, instructions must be provided on how to use these services for work tasks.

It is recommended that an organisation acquires e.g. a licence for a service that it considers safe and provides instructions on what kind of data can be translated using the service in question.

Do not enter any confidential, personal, or otherwise unique data into translation services in your free time wither unless you can ensure the security of the service.

The latest services will communicate with you naturally in a human-like manner, even by changing voices to denote feelings or changing directly from one language to another.

1.3 The services can create and produce articles, emails and almost any other forms of content

AI services are extremely fast and efficient in creating and producing new text content based on the vast amount of information they have been taught with. The significance

of teaching data is emphasised because it affects the content produced by the service, and distortions and inaccuracies can guide the produced data in the wrong direction without the user noticing it.

The organisation must understand the threats associated with using these services and provide instructions and advice on how to ensure the accuracy of the data produced by the services.

Identifying hallucinations or other errors in e.g. images, videos or music may be more challenging than those in text results.

1.4 Services can create music, images and video content

AI services bring significant new opportunities for people who have not previously had the opportunity to produce such content. When will we get to hear the first global hit song created by AI and when will it also be performed by a figure created by AI?

Any copyright issues regarding the text must be taken into account, as with any other material produced.

Artificial intelligence provides cybercriminals and other abusers an opportunity to create completely new types of digital scams. If so far different deep fakes (voice, videos, photographs) have been slow to manufacture and require slightly more advanced expertise, new AI services will make this easy for anyone, with minimal effort.

In the future, competence related to information security and determining the accuracy of information will become increasingly important civic skills. In the future, we will probably have to think more about how we can confirm the authenticity of some public critical information, such as a published video, with a certificate attached to it.

1.5 Change in work tasks - programming as an example

In many contexts, the professions that are expected to be affected by the new AI services have been highlighted. Examples of changes to different professions can be found in the document "GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models”[8]. App development and programming have been highlighted as examples.

Every technological revolution, which the rapidly advancing development of AI services can also be regarded as, has had a wide impact on work. In the early stages, new AI services will increase productivity in several industries and professions by making it possible to do "more in a shorter time". The same applies to such things as application development, but as already mentioned above, organisations must create clear guidelines and policies for the use of AI services in application development, both in its own organisation and in its subcontracting network.

[8] 2303.10130.pdf (arxiv.org)

The incorrect use of AI in application development may cause the organisation's confidential data, code, documentation, algorithms, and production or testing data to fall into the wrong hands. Someone must also check the functionality of the generated code. Even when a code looks like it works, is it otherwise of high quality, effective and appropriate for the purpose of the organisation? The utilisation of artificial intelligence may weaken the organisation's own capabilities and competence and create new types of dependencies on AI services.

As AI services are based on training data, its bias and errors may also be reflected in the code produced.

In the future, we will also need new kinds of professionals who will be able to assess the correctness of decisions produced by different algorithms or services. If it is now important to assess the information security of services, we will also need expertise in the evaluation and tracing of data or decisions produced by the operating logic and service.

1.6 Connecting AI to existing services by using APIs

A conversational, interactive AI service will already provide many users with new opportunities for using it. Another significant, still developing opportunity is linking AI to existing services through APIs. This would allow users to also use the AI service's user interface to utilise the data in the service connected to it. Soon, all services will be available to AI if the service provider wants to enable this.

Would it not be convenient to ask a service to search for the cheapest flight or ferry trip from location A to B, to provide suggestions for four-star hotels in the designated area, to recommend a car rental company or make a travel programme with schedules for sights at the destination e.g. for a week's trip, and recommend local, affordable lunch restaurants with a good reviews?

All of this can be done now, but with several different services and search functions, which can take a significant amount of time. The aforementioned task can already be carried out with the help of AI services when using suitable APIs related to travel services.