SECURE PRACTICE

Exit report from the Secure Practice sandbox project

Topics: legality, Data Protection Impact Assessment (DPIA), fairness, data controller, trust and transparency.

February 2022

Datatilsynet
The Norwegian Data Protection Authority

—

Contents

1. Summary

Purpose of the sandbox project

Allowing yourself to be profiled can make life simpler and more interesting. It's pleasant when streaming services get their suggestions right. It's also undoubtedly more motivating if a course is tailored to exactly your level of knowledge and interests. Profiling can have major advantages both on a personal and societal level However, in the digital age, seeking a more personalised life is a double-edged sword. The more precise the personalisation is, the more precise the personal data is about you, with a risk of abuse.

Profiling in the workplace can be particularly challenging since the relationship between employees and employers has an inherent power imbalance. Employees may find it invasive and insulting. They may feel they are being monitored and fear misuse of the information.

But is there a method that exploits the advantages of profiling while reducing or removing the drawbacks?

This is the starting point for this sandbox project, which examines a new service Secure Practice wants to offer the information security market. The service will use artificial intelligence (AI) to provide individual and personalised security training to employees in clients' businesses.

People are unique, and security training is often too general to be effective. But with artificial intelligence, Secure Practice can offer personalised and therefore more pedagogical training. Both the business and the employees will benefit from better and more interesting training as well as from avoiding fraud and hacking.

Another purpose of the service is for the business to get an overview at a statistical level of knowledge and risk in order to prioritise better measures. The drawback is that individual employees could potentially perceive mapping as invasive. The sandbox project concerns how such a service can be made privacy-friendly.

Conclusions

- **Processing responsibility:** Employers, Secure Practice and the parties have a joint responsibility to comply with the privacy policy. When the AI tool is used in the companies, the employers are the initial data controllers of the processing. When Secure Practice withholds information about which employee the tool is profiling, the employer and Secure Practice are joint data controllers. When the AI tool is developed further during the learning phase, Secure Practice is the sole data controller.
- **Legality:** It is possible to use and develop the service within both general privacy regulations in the EU and special regulations on privacy in working life in Norway.
- **The data subject's rights and freedoms:** With innovative technology and the goal of predicting personal interests and behaviour, it was decided to carry out a Data Protection Impact Assessment (DPIA) for the project. Ultimately the assessment showed that the service has a low risk of discrimination.
- **Transparency:** The project has assessed whether there is a legal obligation to explain the underlying logic in the solution. The conclusion is that there is no legal obligation in this specific case. The sandbox nevertheless recommends more transparency than what is legally required.

Next steps

This final report is important for clarifying the ban on monitoring in the Email Monitoring Regulation.[1] The Norwegian Data Protection Authority will share experiences from the project with the Ministry of Labour and Social Inclusion which is responsible for the Regulation and give the Ministry the opportunity to clarify the rules.

[1] Regulation on Employer's Access to Emails and Other Electronically Stored Material, 2 July No. 1108. The regulation was enacted by the Ministry of Labour and Social Inclusion pursuant to the Working Environment Act.

2. About the project

Secure Practice is a Norwegian technology company that focuses on the human aspect of data security work. Their cloud services are used by over 500 companies, with end users in more than 30 countries. One of the services they currently offer is MailRisk, which helps employees to recognise whether suspicious emails are dangerous or safe by using artificial intelligence. Secure Practice also develops and supplies integrated services for e-learning and simulated phishing.[2]

Now Secure Practice wants to use artificial intelligence to provide personalised security training to employees at their clients' companies. Starting with which interests and knowledge each employee has about data security enables training to be more targeted and pedagogical, and therefore more effective. This tool will also provide companies with reports with aggregated statistics on employees' knowledge and interest level in data security. The reports will make it possible for the employer to monitor development over time, and at the same time identify specific risk areas and uncover any need for collective measures.

Profiling

Profiling is defined in the General Data Protection Regulation (GDPR) as:

"any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements".

In order to provide personalised training, Secure Practice will collect and collate relevant data on the employees at the client's company. The profiling will place each end user in one of several "risk categories", which will determine what training he or she will receive in future trainings.. Risk will be recalculated continuously and automatically so that employees can be moved to a new category when the underlying data dictates this.

The development of the tool is based on a range of scientific studies related to human safety attitudes. Based on these studies, Secure Practice has identified some factors among employees that need to be mapped. Attention has been focused on developing a flexible statistical model and technological solution for processing and linking various data in multiple dimensions, including time. With this as a starting point, the risk assessment itself can be done equally as flexibly, based on which hypotheses form the basis in the model at any given time. These are assumptions that are thus programmed in advance, and the "intelligence" is in the first instance a product of the quality of the hypotheses.

It will also be interesting to be able to use machine learning on data from a historical perspective. So-called learning can be done on the basis of usage metrics to identify patterns for improvement, or possibly deterioration.[3] This will then be able to contribute to improving the hypotheses and developing even more accurate measures and recommendations in the service in the future.

Secure Practice has been working on the new service since Innovation Norway granted funding for the project in 2020. The Research Council of Norway has also granted funding to further develop the theories behind the tool, through a research project with the Norwegian University of Science and Technology. At the launch of the sandbox project, Secure Practice had a theoretical model in place, and technical implementation of the key risk model was within reach. And because the new service is integrated into the existing service platform, a lot of the tool is already complete. At the same time, Secure Practice has still had a number of open-ended questions about both data collection and user interface to consider.

[2] Phishing is a form of social manipulation, where an attacker tries to trick someone into carrying out an action, for example opening a harmful attachment in an email. Simulated phishing involves exercises to help people learn to recognise phishing attempts. Read more at datatilsynet.no.

[3] Read more about learning in point 4.1.1 and 4.1.2

3. Goals of the sandbox process

The sandbox process is divided into three sub-goals, each with their own deliveries. The three sub-goals have been organised thematically around the three central roles that arise when the service is used: the employers, Secure Practice and the employees.

1. **Data controller. Who is responsible for complying with the privacy policy?**
   Is it the operator of the service, i.e. Secure Practice, the company that introduces the service to the workplace or are they both joint data controllers? Is the answer the same during the usage phase as when further development of the tool takes place during the learning phase?

2. **Legality. Can the tool be used and further developed legally?**
   The project will clarify what legal basis the data controllers may have to profile employees in order to offer individually adapted security training for companies and statistical reports to companies. The project will also consider whether such profiling is subject to the prohibition against monitoring in the Email Monitoring Regulation.

3. **The data subject. How does the tool affect the employees?**
   The project will clarify how the data subject is affected with respect to what data forms the basis for the processing, the risks of such processing, fairness, transparency and procedures for exercising the data subject's rights.

4. Assessments and conclusions

4.1 Who is responsible for complying with privacy regulations?

Processing responsibility

The GDPR uses the terms *controller* in Article 4(7), *processor* in Article 4(8) and *joint controllers* in Article 26 to allocate responsibility for complying with the regulations. The accountability principle indicates that the main responsibility for ensuring that the processing of personal data is in accordance with GDPR rests with the controller.

A controller is the person who determines the purposes and means of the processing of personal data, while a processor processes personal data on behalf of the controller. Joint processing responsibility occurs where the parties jointly determine the purposes and means of the processing of personal data.

4.1.1 Assessments of the various roles in the usage phase

A question that emerged early on in the sandbox project was how responsibility should be divided between Secure Practice and their clients. The background to the question was that Secure Practice themselves entered the project with a desire to avoid the client (i.e. the employer) gaining access to individual risk profiles of employees, as a privacy measure to safeguard employees. Secure Practice was nevertheless in doubt about whether it was possible to guarantee this in practice, even if they could implement technical measures in the solution to prevent the client from gaining access to such data.

In other services, Secure Practice has examined its own role as a processor and drawn up ordinary data processing agreements with the client. The client will then be the data controller and decides how the personal data is going to be used. In its role as a processor, Secure Practice is obliged to disclose all personal data the client wants to gain access to. An inherent risk of such a division of responsibility in the new service is that Secure Practice cannot decline to disclose personal data on individual employees to the employers, even in the event of suspicion that the data will be used for other purposes (for example to influence salary levels or bonus payments).

In collaboration with Secure Practice, the Norwegian Data Protection Authority researched the consequences of various divisions of responsibility, and an alternative solution involving joint processing responsibility was suggested. Secure Practice wishes to withhold personal data from their clients. By withholding data from their clients they have a crucial influence on the processing of this personal data which extends beyond their role as processor.[4] This means that Secure Practice and the client jointly determines the purposes and means of the processing of the employees' personal data.

In its guidelines, the European Data Protection Board refers to a distinction between "essential" and "non-essential" means. Section 40 of the guidelines links "essential" means to the choices that are transferred to the data controller. As stated in the guidelines there is a close connection between what constitutes "essential means" and the issue of whether the processing is legal, necessary and proportionate. "Non-essential" means are linked to practical implementation, for example security measures.

To minimise the employer's access to the employee's personal data can be seen as a means to reducing privacy disadvantages. In the assessment of the legal basis and therefore of whether or not the tool is legal, the privacy disadvantages are relevant, see point 4.2 below.

European Court of Justice case law shows that parties can become joint data controllers, even if the processing of personal data is not evenly divided between the parties or the client does not have access to the personal data that is being processed. This might be the case where the service processes the personal data for its own purposes, and this processing can be carried out only because the client facilitates such processing by choosing the service.

According to the way the service is outlined in the sandbox project, Secure Practice's processing of personal data will only be possible because the client uses the service. Secure Practice processes this personal data by preventing personal data on each individual employee from being disclosed to the client. Withholding personal data from the client accordingly creates a joint processing responsibility between Secure Practice and their clients.[5]

The purpose of organising the processing with joint processing responsibility is to ensure that the division of responsibility reflects the actual role that Secure Practice undertakes in the individual processing situations. This means that there is no change in the processes where Secure Practice actually acts as a processor and processes personal data on behalf of the client. However, where the processing is not exclusively done on behalf of the client, the GDPR requires the parties to identify this.

Secure Practice and the individual client must map the processes where they jointly determine the purposes and means of the processing so that they determine responsibility among themselves in a transparent manner. A transparent determination of responsibility between Secure Practice and their clients is intended to prevent the diffusion of responsibility between the companies when employees seek to exercise their rights under GDPR.

The allocation of responsibility can take place through a contract or other document between the client and Secure Practice. Regardless of how this is arranged between Secure Practice and the client it must be communicated outwardly so that employees are aware of how they can request access to their personal data and exercise their rights in accordance with GDPR.

[4] EDPB, Guidelines 07/2020 on the concepts of controller and processor in the GDPR, section 40.

[5] The European Court of Justice has argued that concurrent interests argue for joint processing responsibility in the Judgment of 29 July 2019, *Fashion ID* C-40/17, section 80. "As to the purposes of those operations involving the processing of personal data, it appears that Fashion ID's embedding of the Facebook 'Like' button on its website allows it to optimise the publicity of its goods by making them more visible on the social network Facebook when a visitor to its website clicks on that button. The reason why Fashion ID seems to have consented, at least implicitly, to the collection and disclosure by transmission of the personal data of visitors to its website by embedding such a plugin on that website is in order to benefit from the commercial advantage consisting in increased publicity for its goods; those processing operations are performed in the economic interests of both Fashion ID and Facebook Ireland, for whom the fact that it can use those data for its own commercial purposes is the consideration for the benefit to Fashion ID." (Our underlining).

ℹ Artificial intelligence in three phases:

We often divide artificial intelligence into three phases: *development, usage* and *learning*. The question of data protection arises in all phases.

The development phase: Machine learning is sometimes used to gain insight into large data quantities, insight that can in turn be used to develop new solutions. For example, Secure Practice has carried out an analysis of emails to find out what characteristics are particularly present in *suspicious* emails. These findings could then be turned into artificial intelligence to detect harmful emails, through specific patterns that developers incorporate into the software.

Usage phase: Machine learning can also be used to discover new connections in real time, without humans needing to be involved in the process. In other words, the degree of automation here will be high, unlike the development phase, although it may otherwise be very similar. Secure Practice already uses machine learning in real time in its MailRisk service, without humans being involved, to identify and classify hitherto unseen scams and cyber-attacks that spam filters have been unable to stop.

Learning phase: This is more similar to the development phase, but generally with more data available, particularly data on the results after the system has done its job. Learning can be crucial for verifying or improving correctness in a machine learning model. This applies in particular to conditions that change over time, such as fraud methods used in malicious emails. Learning is also relevant for complex conditions, such as safety attitudes among the company's employees.

Secure Practice has not used personal data in the development phase of the service to which the project applies. It has therefore been decided to focus this project on the usage phase which illustrates the specific challenges in the workplace most clearly. We have also briefly discussed learning at various points, to illustrate that machine learning is a continuous process.

4.1.2 Assessments of the various roles in the learning phase

The use of personal data for learning to improve one’s own products is, unlike the other processing, primarily of interest to Secure Practice, although their customers may also benefit from the results of such learning.

Secure Practice determines the purposes and means of this processing. The learning could therefore not be carried out without the personal data used in the service.

To ensure that Secure Practice has the sole processing responsibility for the learning phase, controls should be established in the user interface so that the client can prevent the use of personal data for this purpose. This separation ensures that there is no lack of clarity about when Secure Practice is acting as a controller and when the company is acting as a data processor. Furthermore, the information on the processing should be clearly presented to clients who wish to use this functionality.

4.2 Can the AI tool be used and further developed legally?

In this chapter we look at the legal basis for the solution, and at the national prohibition against monitoring in the Email Monitoring Regulation.

4.2.1 What rules apply?

Anyone who wants to collect, store or otherwise process personal data must comply with the requirements in the General Data Protection Regulation (GDPR). In the following we will examine more closely the fundamental requirement that there must be a legal basis (consent, agreement, basis in law etc.) for processing personal data.

The GDPR applies with equal force in 30 countries in Europe. Norway also has special rules on privacy in working life. These special rules are provided in regulations enacted pursuant to the Working Environment Act.[6] The regulation relevant for Secure Practice is the Regulation of 2 July 2018 on Employer's Access to Emails and Other Electronically Stored Material, commonly referred to as the "Email Monitoring Regulation". The Email Monitoring Regulation concerns employers' access to all electronic equipment employees use in a work context and is intended to protect employees against unnecessarily invasive monitoring or control. Section 2, second paragraph of the Email Monitoring Regulation prohibits the monitoring of employees' use of electronic equipment. In the sandbox project we have assessed how this affects the tool Secure Practice is developing.

4.2.2 Usage phase

Before Secure Practice offers the tool on the market, it is important to find out whether future clients will be permitted to use the tool. In the first workshop in the sandbox project, we started by looking at whether an employer that purchases the service has a legal basis for processing personal data from the employees in this tool. We also examined more closely how the service is affected by the employer's prohibition against monitoring employees, as provided for in the Email Monitoring Regulation.

We note briefly that for the processes where Secure Practice and the employer have joint processing responsibility, both must have a legal basis for the processing. In this chapter on the usage phase, we have chosen to focus on the employer as the data controller.

What personal data was relevant in the AI tool and was assessed in the sandbox project?

Before we assessed the legal basis, Secure Practice listed all possible data that may be valuable for mapping employees' interests and knowledge within information security. The purpose of the exercise was to assess more closely, what data is relevant and decide what is acceptable to use from a privacy perspective.

Data Secure Practice assessed to be both desirable and in the "right" end of the privacy scale was data on the completion of e-learning, phishing exercises and reporting from the MailRisk service, as well as answers to surveys and quizzes about knowledge and habits related to information security.

Then data that might be borderline was listed, with regard to both privacy and usefulness. These were the most important to discuss, for example demographic data such as *age* and *seniority* of the employees, and data from other security systems such as weblogs.

Finally came data that was never actually relevant from a privacy perspective to use, but which was nevertheless included in the original list due to potential usefulness. Examples in this category include psychological tests or data collected from social network profiles.

[6] Act relating to working environment, working hours and employment protection, etc. of 17 June 2005

Brief explanation of the relevant legal basis – legitimate interests

Secure Practice and the Norwegian Data Protection Authority assessed that “legitimate interests” in accordance with GDPR Article 6(1)(f) was the most relevant legal basis in this project.[7]

The provision states three conditions which must all be met for processing to be legal:

1. The purpose of the processing must be linked to a legitimate interest
2. The processing must be necessary to achieve the purpose
3. The employees’ interests, rights and freedoms must not exceed the employer’s interests. In short, we call this step “balancing of interests".

Condition No. 1 - legitimate interest

As noted above, the purposes of using the tool are two-fold:

1) To give employees individually adapted training in information security.

2) To provide the companies with statistical reports that describe employees’ knowledge and interest levels in information security at a group level.

Both these purposes are linked to the company's interest in improving information security. We assessed better information security to be in the interests of the company itself, the employees, third parties such as clients and partners and society as a whole. It was easy to conclude that improvement of information security constitutes a legitimate interest and that the first condition is therefore met. The discussions in the project therefore concerned the last two conditions.

Condition No. 2 - necessity

Useful questions when assessing which processing is necessary:

- Will the processing of this personal data actually help achieve the purposes?
- Can the purposes be achieved without processing this personal information or by processing fewer personal data?

The employees’ knowledge and interest in information security must be mapped to achieve both purposes – both individually adapted security training and statistical reporting for the companies. The discussions in the sandbox concerned how Secure Practice can minimise the use of personal data and ensure that the data used actually helps to achieve the purposes.

We formed focus groups of employees from the trade union Negotia and a large company in order to gain potential users’ perspective on the assessments. Interesting insight was provided into the assessment of how apparently appropriate data can be the exact opposite.

One of the methods Secure Practice wants to use for the mapping is surveys. In these, employees must take a position on various statements, such as “I have deliberately broken information security regulations at work”. The focus group of representatives from the trade union commented that it is unclear what consequences there could be for the individual employee to answer such a question if they had actually broken the rules. In the second focus group emphasis was placed on the importance of anonymity towards the employer if one were to give an honest answer to this.

Regarding the purpose of individually adapted training in information safety, this question will not help to achieve the purpose if employees who have broken security rules answer no because they are afraid of the consequences of answering yes.

As regards to the purpose of creating statistical reports for the companies, the question viewed in isolation could contribute to achieving the purpose as long as some of the employees would answer yes. Such feedback could suggest to the company that the

[7] Read more about the provision on the Norwegian Data Protection Authority’s website on the legal basis, and on the British Information Commissioner’s Office’s website What is the ‘legitimate interests’ basis?

internal regulations on information security are badly designed or do not fit in to the everyday working life of the employees. The example illustrates that the data controller must assess the legal basis separately for each purpose the processing is going to address.

As stated in point 4.4 of this report, Secure Practice was advised to reformulate some of the questions in order to more easily achieve honest answers.

Once Secure Practice has identified the personal data necessary for the tool to work for individual training and overall information to the employers, the next step will be a balancing of interests.

Condition No. 3 - balancing of interests

The third condition concerns the employer being unable to introduce measures if the employee's interests, rights and freedoms carry more weight than the employer's interests. In short, a balancing of interests is about finding a balance between the interests on both sides, so that any invasion of privacy is proportionate. In order to carry out this balancing of interests, we started by investigating how the employees are affected by the tool.

Employees are in an uneven power relationship with the employer. It is therefore especially important for employees that personal data about them is not misused for new purposes. An employee who is "tricked" during a phishing exercise expects the result only to be used for training purposes and not to assess what assignments or benefits he or she receives. But if the employer gains access to this information, there is a risk of such misuse.

A distinct feature of solutions that use artificial intelligence is that they often process large quantities of personal data.[8] The tool we are discussing in this project is meant to map both knowledge and interest and is suited for a detailed survey of employees. This can be perceived as invasive. The solutions within artificial intelligence can also produce an unexpected result.

We assess consideration for the employees to carry a great deal of weight, and greater demands are therefore placed on the interests linked to the employer. On the other hand, the interest linked to better information security also carries a great deal of weight. As mentioned in the point on legitimate interest, better information security is an interest that benefits the individual employees, not solely the employer.

When balancing the interests linked to information security and consideration for the employees' privacy, we assessed these points in particular:

- How the employees will perceive the mapping and what consequences it might have for them. Positive consequences may be that they receive personalised help and follow-up to strengthen their skills in information security, and that they avoid being affected by fraud and hacking with the consequences this may entail. Potential negative consequences may be that employees feel unsure about how data about them will be used, and whether there may be negative consequences if they reveal little knowledge or interest in information security.
- How the employers involve the employees before they introduce the tool.
- What information about each employee the employer has access to.
- How Secure Practice provides information to employees in the tool and in the privacy policy.
- What technical guarantees are built in to prevent outsiders from gaining access to personal data on the employees.

As regards to the bullet points on possible consequences for the employees and what information about individual employees the employer has access to, Secure Practice worked on the assumption that the employer should not have access through the tool to measurements of each individual. The sandbox's recommendation is the introduction of both *legal* and *technical* guarantees that information about the employees does not go astray.

[8] Any discrimination in the tool will also have a major impact since this affects the individual's career. As regards to possible discrimination in the solution, please see the assessment from the Gender Equality and Anti-Discrimination Ombud below stating that the risk of discrimination is low (see point 4.5). The risk of discrimination has therefore not been emphasised in the balancing of interests.

By legal guarantees we mean that Secure Practice should include provisions in the contract that employers do not have access to information on individual employees – not even upon special request. By technical guarantees we refer to the measures Secure Practice has already implemented to prevent the employer or others from gaining access to this information. The service uses both pseudonymisation and encryption to protect personal data. The name of the employee is replaced with a unique identifier, and the link between the name and the identifier is stored in a separate database. Secure Practice has also identified a need to store the user's email address in proximity to the user's other information. A personal email address establishes a clear link to an individual and will therefore challenge the original objective of a pseudonymised database.

In order to accommodate this challenge, Secure Practice has chosen to encrypt email addresses; names and other direct identifiers in the database itself, so that these will no longer be available in the open. The keys used to encrypt and decrypt such identifiers are stored separately from the database. In this way anyone who gains access to the user database will not gain access to the information necessary to link personal data to individuals.

When both legal and technical guarantees are in place, the sandbox assesses the risk of the employer or others being able to access and possibly misuse personal data that the tool collects to be low. This contributes to the overall interests carrying most weight on the part of the data controller. It will therefore be possible to use legitimate interest as a legal basis in the usage phase.

The right to object

When legitimate interest is used as a legal basis, employees have the right to object to the processing of personal data in accordance with Article 21 of the GDPR. If an employee objects to the use of their personal data in the tool, the employers must take into consideration the specific circumstances the employee has pointed to in the objection. The employer must still carry out a balancing of interests with respect to the person who has protested, but must show that there are "*legitimate compelling reasons*" for using the personal data in the tool.[9]

Accordingly, the assessment must be done on an individual basis, taking account of the justification the employee who is objecting has given. A data controller assessing an objection must to a greater extent consider alternatives to specifically adapted training and reporting at a statistical level. Secure Practice envisages that those who object will either have their objection processed by the employer, or have it granted through the tool without further manual processing, as it may be challenging for Secure Practice to assess the basis for the objection.

If multiple employees object and do not use the tool, a smaller proportion of the company will understandably be mapped by the tool. Data controllers must be aware that this may make it easier to identify the employees in the reporting at a statistical level.

4.2.3 Learning

Use of personal data for learning to improve the tool also requires a legal basis.[10] Secure Practice is the sole data controller for this phase. In the same way as for the usage phase, the purposes of improving the service are linked to a legitimate interest. Learning is expected to gradually make the service more accurate as personal data from more employees is entered into the tool. This will likely increase information security in clients' companies.

Unlike the usage phase, however, the purposes are still clearly linked to commercial interests, since an expected increase in quality may offer increased sales. We would therefore briefly note that commercial interests are also legitimate interests, and accordingly the first condition is met.

As regards to the second condition, Secure Practice must take an active position on what personal data is necessary to achieve each individual purpose. We assume that the same type of personal data as in the usage phase is relevant to making the service more accurate. If Secure Practice is to test the tool for possible discrimination, further

[9] See Article 21 GDPR.

[10] See point 4.1 for a more detailed explanation of learning. In the sandbox project, it was limited to issues related to the reuse of personal data and restrictions that follow from the purpose limitation principle in Article 5(1)(b) and Article 6(4). The French Data Protection Agency has published guidance on the use of personal data for learning on its website (cnil.fr) under the title "Sous-traitants: la réutilisation de données confiées par un responsable de traitement".

assessment of what information is necessary . For this purpose, access to for example demographic data such as age and gender may be highly useful. However, we have not examined this assessment in this project.

When balancing consideration for Secure Practice's interests and employees' privacy, Secure Practice may emphasise that increased accuracy will benefit both the companies and the employees. If the solution is not updated through learning, the tool may become outdated and not function as intended. Possible privacy disadvantages for the employees may be greater during this phase, as personal data will be used to further develop the tool for *new* companies. With the legal and technical guarantees discussed under the usage phase in place, it will in our opinion also be possible to use legitimate interest as a legal basis in the learning phase.

In this phase too, employees can object against processing.

4.2.4 The national prohibition against monitoring in the Email Monitoring Regulation

So far, we have focused on the GDPR. However, it is also relevant to assess the service in relation to the Email Monitoring Regulation with its prohibition against "monitoring employees' use of electronic equipment, including use of the internet". In contrast to the GDPR, it takes more than the processing of personal data for these special rules to apply.

Another important distinction from the GDPR is when an eventual monitoring is legal. There are two legal instances. Either to "manage the company's computer network" or to "detect or resolve security breaches in the network".

In other words, an employer can lawfully use the tool as long as the use does not entail monitoring of the employees, or if one of the two instances named above are met. But can the use of the AI tool be seen as monitoring of the employees' use of electronic equipment?

What counts as "monitoring" is not defined in more detail in the Regulation. The legislative history of similar regulations in the previous act highlights that the measure must have a certain duration or take place repeatedly.[11] Monitoring is in contrast to individual access which is permitted in several situations. The legislative history also emphasises that it is not solely a question of whether the purpose is to monitor. The employer must also attach weight to the question of whether the employee may perceive the situation as monitoring.

The AI tool Secure Practice is developing is composed of many methods for mapping the employees. The methods range from surveys and quizzes to recording how employees react to simulated phishing exercises and activity in the learning platform. Viewed in isolation, this would not count as monitoring the use of electronic equipment. However, the question is whether the overall mapping is affected by the prohibition against monitoring.

Past decisions by the Norwegian Data Protection Authority are not conclusive with regard to whether the employer is required to actually see the data or metadata for it to count as monitoring. The concept of monitoring is broad and it may be the case that collection and systematisation are also affected by the prohibition. The fact that the provision is targeted at the *employer's* monitoring indicates that the employer must at the very least be able to access the data on the employees in order to be subject to the prohibition.

After discussions in the sandbox, there was a consensus that the mapping in the tool would not be affected by the monitoring prohibition. We have specifically emphasised the technical and legal measures that have been implemented to ensure that the employer will not have access to the data that is collected on each employee.

The statistical reports for the employers concerning the level of information security among the company's employees may be more likely to be perceived as monitoring. The reporting must take place at the group level. The number of employees the company has and the number of staff in each group will likely affect how the employees perceive the tool. We have assumed that the data

[11] The Ministry of Government Administration, Reform and Church Affairs consultation paper: Proposal for regulations on employer's access to employees' emails etc., 17.10.2006, p. (15). "Monitoring" means that the measure must have a certain duration or must take place repeatedly. The inverse is the situational access which is permitted under the regulations here.
The provision covers both automatic and manual monitoring. The provision must be seen in conjunction with Sections 7-11, last paragraph, which prohibits the use of activity logs for controlling and monitoring individuals." (Quote from Signhild Blekastad and Marion Holthe Hirst: Privacy and control in working life p. 310).

is provided in a way that does not enable the employer to identify individual employees. What information the employees receive will also influence how they perceive the service.

The sandbox has concluded that the mapping, which is only communicated to employees, is not affected by the prohibition against monitoring. As regards to measuring the security level of the employees at group level, the employer must consider the design in more detail. We recommend that the employer discusses how the reporting may take place with the employees beforehand.

Since the use of the tool does not count as monitoring in this case, we do not need to assess the rest of the provision.

4.2.5 Which (related) regulations have not been assessed in the sandbox project?

We have not assessed whether the employer's use of the AI tool falls under the regulations on control measures in the Working Environment Act, Chapter IX. The Norwegian Labour Inspection Authority enforces these regulations, and you can read more on their website.

We have also not examined the regulations that apply to *access* to information stored on a user's terminal equipment, computer, telephone etc. as regulated in the Electronic Communications Act.

4.3 Data Protection Impact Assessment – DPIA

If it is likely that a type of processing of personal data will entail a high risk to people's rights and freedoms, the data controller must assess the consequences for privacy of the planned processing. This particularly applies when using new technology.

Determining with certainty if there is a high risk can be challenging. If there is uncertainty about this issue, the Norwegian Data Protection Authority recommends carrying out a Data Protection Impact Assessment (DPIA). This can be a useful tool for ensuring that the other requirements in the GDPR are met.

The Norwegian Data Protection Authority has produced a list of processing activities that always require a DPIA to be carried out.[12] From this list the following elements are relevant to the tool Secure Practice is developing:

- Processing of personal data with AI which is innovative technology.
- Processing of personal data for systematic monitoring of employees.
- Processing of personal data where the purpose is to offer a service or develop products for commercial use which involves predicting job performance, finances, health, personal preference or interests, reliability, behaviour, location or patterns of movement. (Specific categories of personal data or highly personal data and evaluation/scoring).

The sandbox therefore concluded that the use of Secure Practice's tool requires a DPIA. It is the responsibility of the controller to ensure that a DPIA is conducted. In practice, this means that companies that purchase the new service from Secure Practice must also carry out a DPIA.

For many small and mid-size companies it can be demanding to implement a sufficient DPIA of a tool that is based on artificial intelligence. This requires, inter alia, knowledge of privacy regulations and other basic rights, artificial intelligence and knowledge of the system's logic in addition to the particular conditions in each workplace.

[12] Can be found at datatilsynet.no under the title "Når må man gjennomføre en vurdering av personvernkonsekvenser". The list is based on guidelines from the Article 29 Working Party, Guidelines on Data Protection Impact Assessment (DPIA) (wp248). The guidelines have been approved by the European Data Protection Board.

The asymmetrical relationship between the client and the service provider is often evident in our digital society. To exaggerate slightly, it may occasionally seem like the supplier sets requirements for the client instead of the other way around. An equivalent dynamic can also be found between an AI services provider and the client. In this situation the provider will also have the role of technical expert which can highlight both advantages and disadvantages of the technology they sell.

The Norwegian Data Protection Authority and Secure Practice agreed that responsible use of AI requires a data controller with a solid information basis, which enables the data controller to carry out the correct assessments. Based on this the sandbox decided to include a DPIA in the project.

It is important to stress that it is not sufficient to assess the data protection impacts of the tool itself. The assessment must also take account of the context the tool will be used in. This context will often vary from client to client, and which part of the country (or the world) the client lives in. This means that Secure Practice can undertake part of the investigative work for the client in advance, but the assessments for each specific circumstance must be performed by the client themselves.

It was important for the Norwegian Data Protection Authority to be able to provide effective guidance in the process for assessing data protection impacts, without the advice leaving so little room for manoeuvre that Secure Practice's ownership of the process was challenged. This was particularly important as the development of the service was in an early phase when the Norwegian Data Protection Authority gave feedback. Secure Practice themselves arranged a workshop for assessing data protection impacts together with the Norwegian Data Protection Authority, and then documented the results.

The Norwegian Data Protection Authority gave feedback on various issues linked to potential consequences for data protection and the design of the assessment itself:

- Publishing the DPIA online can be one of several measures to facilitate transparent processing of personal data. Publishing the assessment is not in itself enough to fulfil the duty to inform. The data subject must be informed in a concise, transparent, comprehensible and easily accessible manner.
- It is important to ensure procedures so that the privacy policy and the DPIA are updated in parallel. This means that changes and new solutions that are implemented in the production solution must be reflected and dealt with in the DPIA and in the privacy policy when this is relevant.
- It is important to avoid ambiguous and sweeping wording. The descriptions should be as precise as possible so that it is possible to see what has been assessed. The object of assessment must be clearly stated.
- Messages directed at the client (company) must avoid wording that may be confused with the legal basis for the processing of the employee's personal data. This was particularly relevant where the client's agreement with Secure Practice was mentioned, and this could be confused with the reference to Article 6(1)(b) GDPR ("processing is necessary for the performance of a contract to which the data subject is party [...]").
- Threshold values for consequence and probability must take particular account of the risk for each data subject, and that the DPIA does not exclusively concern how many people are affected, but also the consequences for each data subject.
- The Norwegian Data Protection Authority recommended that the risk to the data subject's freedoms and rights be investigated in more detail with regard to the intended division of responsibility between Secure Practice and the company. This can clarify both the division of roles and responsibilities for the data subject.

Feedback on the DPIA above is focused on the usage phase. Finally, we mention that Secure Practice must also assess the data protection impact for the learning phase where they have independent processing responsibility.

4.4 How to explain the use of artificial intelligence

Processing personal data in a transparent way is a fundamental principle of the Personal Data Act. Transparency enables the data subject to exercise their rights and safeguard their interests. In the sandbox project we discussed what requirements there are to inform the data subject about how personal data is processed. We also discussed specific issues relating to the user interface.

4.4.1 What requirements are set for transparency?

The requirement to provide information to the data subject can be found in Articles 13 to 15 in GDPR. Article 13 regulates what information is to be provided when obtaining personal data from the data subject. Article 14 regulates what information is to be provided, and when this information is to be provided, if the personal data is not obtained from the data subject themselves. Article 15 regulates the data subject's right of access to personal data about them that is being processed. Article 12 also imposes a general obligation to provide information in a concise, transparent, intelligible and easily accessible form, using clear and plain language.

Regardless of whether or not you use artificial intelligence there are certain requirements for transparency when processing personal data. In summary these are:

- The data subjects must receive information on how the data is used, depending on whether the data is obtained from *the data subject* themselves or from others.
- The information must be easily accessible, for example on a website and be written in clear and intelligible language.
- The data subject has the right to know whether data about them is being processed and have access to their own data.
- It is a fundamental requirement that all processing of personal data must be done in a transparent manner. This means that an assessment must be carried out of what transparency measures are necessary for the data subjects to be able to safeguard their own rights.

In the first bullet point there is a requirement to provide information on how the data is used. This includes, inter alia, contact information for the data controller, the purpose of the processing and what categories of personal data will be processed. This is information that is typically provided in the privacy policy.

These duties are targeted at the data controller. When Secure Practice and the employer have joint processing responsibility, they must determine which responsibilities each of them has in order to meet the requirements in the GDPR.[13] The obligation to allocate responsibility follows from Article 26 GDPR. This means, inter alia, information on how the data subjects can exercise their rights and what personal data about them will be processed in the tool.

4.4.2 Do the employees have a right to be informed about the logic of the algorithm?

For automated decisions which have a legal effect or significantly affect a person, specific requirements for providing information apply. Article 13(2)(f) states that the data controller in these cases must explain the underlying logic of the algorithm. The same applies in accordance with Article 14(2)(g) when the personal data is not obtained directly from the data subject.

Guidelines from the Article 29 Working Party

"Articles 13(2) (f) and 14(2) (g) require controllers to provide specific, easily accessible information about automated decision-making, based solely on automated processing, including profiling, that produces legal or similarly significant effects.

If the controller is making automated decisions as described in Article 22 (1), they must:

- tell the data subject that they are engaging in this type of activity;
- provide meaningful information about the logic involved; and
- explain the significance and envisaged consequences of the processing"

(Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679, page 25.)

[13] See more about joint processing responsibility in point 4.1. and 4.2.

The tool in this sandbox project neither leads to legal effects for the employees nor affects them significantly. The processing therefore falls outside the scope of Article 22 of the Regulation. The duty to inform in accordance with articles 13 and 14 is aimed at processing covered by Article 22. Accordingly, no duty to explain how the algorithm functions follows directly from this provision.

The project assessed whether the principle of transparency read in light of the recital could imply a legal duty to inform how the algorithm functions.

According to Article 5(1)(a) in GDPR, the data controller must ensure that personal data is processed fairly and transparently. Recital 60 highlights that the principle of transparent processing requires that the data subject be informed of the existence of profiling and the consequences of such profiling. The recital refers to profiling in general, and it therefore appears to be somewhat broader in scope than Article 13(2)(f) and Article 14(2)(g) which refer to automated decisions with legal or other significant consequences.

Recital 60 in the GDPR

"The principles of fair and transparent processing require that the data subject is informed of the existence of the processing operation and its purposes. The controller should provide the data subject with any further information necessary to ensure fair and transparent processing taking into account the specific circumstances and context in which the personal data are processed. Furthermore, the data subject should be informed of the existence of profiling and the consequences of such profiling. Where the personal data are collected from the data subject, the data subject should also be informed whether he or she is obliged to provide the personal data and of the consequences, where he or she does not provide such data. That information may be provided in combination with standardised icons in order to give in an easily visible, intelligible and clearly legible manner, a meaningful overview of the intended processing. Where the icons are presented electronically, they should be machine-readable."

The Article 29 Working Party has given its views on transparency in processing situations which fall outside Articles 13, 14 and 22. In the guidelines concerning transparency, the importance of explaining the consequences of processing personal data is emphasised as well as that the processing of personal data must not come as a surprise to those who are having their personal data processed.[14] The fact that the obligations to explain the underlying logic in accordance with Articles 13 and 14 go beyond the general principle of transparency as discussed in Recital 60 is also supported by the guidance on profiling and automated decisions.[15] In summary, it is difficult to see how a legal obligation to explain the underlying logic of the tool in this project can be deduced from the Regulation corresponding to the requirements of Articles 13 and 14. In any case, the Article 29 Working Party states in the aforementioned guidelines that it is good practice to explain the underlying algorithm, even if the data controller does not have a duty to do so.

The sandbox also recommends providing information on how the tool from Secure Practice works, as this can help create trust in the AI tool. In the section below, we refer to an example from a focus group of employees, who highlighted the importance of clear and plain information as a prerequisite for providing correct personal data.

[14] Article 29 Working Party, Guidelines on transparency under Regulation 2016/679, WP260 rev. 1, section 41. The guidelines have been approved by the European Data Protection Board.

[15] See more on p. 25 of the Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679 (WP251REV0.1) The guidelines have been approved by the Data Protection Board.

4.4.3 How and when is it best to provide information to users?

The GDPR does not regulate in detail how the user interface should be designed. But in continuation of the question of the duty to inform, attention was also focused on *how* and *when* the solution should inform the user.

In the project we discussed, inter alia, specific issues linked to the design of the user interface. An important point was whether employees should receive an explanation of why the AI tool is offering you this exact suggestion, whether you are being encouraged to complete a specific training module or take a specific quiz, and how it should be carried out.

A specific example could be an employee receiving a suggestion to complete a certain type of training because they had been tricked by a phishing exercise. Such information reveals something about the underlying logic of the algorithm. It was specifically discussed whether such detailed information might make the user feel they are being monitored, which could again lead to decreased trust. The arguments in favour of providing this type of information were that the data subjects need it to understand how the data is used and that this understanding can build trust in the solutions.

In the first focus group there was a high degree of willingness to use the solution and share data if it constructively contributed to achieving the goal of better information security in the company. The participants emphasized the importance of clear and plain communication with the employees. It is important to clarify early in the process how the data will be stored and used in the company's work. Uncertainty surrounding how the data will be used increases the danger of the employees adapting their answers to what they believe is “correct” or being unwilling to share data. This is an interesting finding because the algorithm becomes less accurate if the data it is based on is inaccurate and does not represent the actual situation of the user.

In the focus group with the trade union, Negotia, there was a major focus on transparency in general as a prerequisite for employees to be able to trust the solution. The points that were emphasised were linked, inter alia, to what information the employer has access to, how the contract with the company is designed, the importance of involving the employees or union representatives at an early stage in the process and that such a solution may be perceived differently by employees depending on the situation, for example whether they have a high or low degree of trust in the employer. The risk associated with the traceability of answers back to individual employees was also highlighted in the focus group. This focus group warned against designing questions in such a way that the answers could damage the employees if the employer became aware of them.

4.5 Fairness and discrimination

The GDPR places fairness among the fundamental principles governing the processing of personal data in Article 5 and is mentioned along with transparency concerning the data subject and the legality of the processing. A similar distinction is used in the EU expert group’s ethical guidelines for trustworthy AI[16] and is reflected in the National AI Strategy with the key principles *legal*, *ethical* and *secure*.

There is still little practice that can clarify the specific content of the principle of fairness in more detail, but there is some guidance in the recitals of the GDPR. It is also worth mentioning that the principle of fair processing of personal data is meant to be a flexible legal standard which can be adapted to the specific processing situation.

The specific situation of the data subject must be taken into consideration when assessing whether the processing is compatible with the requirements in the Regulation. One must take into what reasonable expectations the data subject has for protection of their personal data, as well as any power imbalance between the data subject and the data controller.

[16] You can find “Ethical guidelines for trustworthy AI” at europa.eu.

In its guidelines of 4/2019[17] on embedded privacy, the European Data Protection Board highlights several aspects included in the principle of fairness, including non-discrimination, the expectations of the data subject, the broader ethical issues of the processing and respect for rights and freedoms. In order to ensure that the solution was also assessed in a broader perspective in terms of fairness, the project involved the Gender Equality and Anti-Discrimination Ombud (LDO).

In a separate workshop, the LDO presented links between privacy regulations and the Equality and Anti-Discrimination Act. The Ombud also gave an introduction to how an actor can assess whether unlawful discrimination is taking place in accordance with Sections 6-9 of the Equality and Anti-Discrimination Act.

The LDO emphasised the following concerning Secure Practice's service:

In an early phase of the project, Secure Practice launched an option where those with the best scores in the tool could work as ambassadors in their company. The LDO pointed out the risk that the ambassadors would have a more positive career trajectory than others if employers had access to each employee's score. The LDO also showed that if the AI tool rewarded characteristics and interests which for example are most common in individual groups, there would be a risk of indirect discrimination in the model. Indirect discrimination is generally unlawful in accordance with Section 8 of the Equality and Anti-Discrimination Act.

Secure Practice is now taking further steps to ensure that employers do not have access to individuals' scores, which the LDO believes is a good measure to reduce the risk of the employer being able to use the information concerning employees' knowledge of cyber security for purposes other than intended. At the same time, the LDO believes that going forward it is important for Secure Practice to specify what demographic data is to be collected, how this data will be used as well as the justification for and factuality of the relevant data use.

The LDO encourages Secure Practice to ensure that the training and course adaptation work equally as well for all groups; women and men, different age groups and people with disabilities, etc. In order to achieve this, Secure Practice should avoid playing on stereotypical ideas about various groups when the training is adapted to the different employees. Playing on stereotypes is not necessarily discrimination but can contribute to reinforcing traditional ideas about people who belong to specific groups. Such ideas can be less accurate which will reduce the value of the training adaptation. Other measures encourage Secure Practice to periodically test, and where necessary, adjust their own service to ensure that the customization features work equally as well for all users.

Under these conditions the LDO believes that the discrimination the employees are exposed to by the training being adapted to the individual's skill level, is an objective form of discrimination, see Sections 6 and 9 of the Equality and Anti-Discrimination Act.

[17] European Data Protection Board Guidelines 4/2019 on Article 25 Data Protection by Design and by Default

5 Next steps

In the sandbox, Secure Practice has explored issues they had in the product development process where built-in privacy is crucial. By taking initiative for this project and remaining at the forefront, Secure Practice can use the results to ensure that the new service they are launching provides effective protection of privacy from the start.

This final report is important for clarifying the prohibition against monitoring in the Email Monitoring Regulation. The Norwegian Data Protection Authority will share experiences from the project with the Ministry of Labour and Social Inclusion which is responsible for the Regulation and give the Ministry the opportunity to clarify the rules.

The Norwegian Data Protection Authority is also working to update the information on privacy in the workplace on the website. More than a quarter of inquiries to the Norwegian Data Protection Authority's guidance services and a large number of complaints concern this topic. In 2021 alone, 1677 questions about privacy in the workplace were received, so this is an important topic to spread awareness about.

Datatilsynet

The Norwegian Data Protection Authority’s regulatory sandbox for responsible artificial intelligence

Office address:
Trelastgata 3, Oslo

Postal address:
PO Box 458 Sentrum
0105 OSLO

sandkasse@datatilsynet.no
Phone: +47 22 39 69 00

datatilsynet.no/sandkasse
personvernbloggen.no
twitter.com/datatilsynet