Universiteit Leiden

Master ICT in Business and the Public Sector

RPG: What role should NCSC-NL play?

Name: Bram van Aggelen
Student ID: 1840045
Date: 29/08/2024

Specialisation: ICT in Business

1st supervisor: Eleftheria Makri

2nd supervisor: Olga Gadyatskaya

Master's Thesis in ICT in Business and the Public Sector

Leiden Institute of Advanced Computer Science
Leiden University
Einsteinweg 55
2333 CC Leiden
The Netherlands

Abstract

Generative Artificial Intelligence (AI) is bringing changes in the field of cyber security. NCSC-NL is concerned the speed and impact of the new capabilities offered by generative AI is making it hard to perform their statutory task. NCSC-NL has identified a research-practice gap where researchers are finding and improving new capabilities, but these are not seen in use by the practitioners. This is a phenomenon has its basis in other types of research which are explored in Section 2. To aid NCSC-NL in bridging this gap a survey and two types of interviews is be conducted with NCS-NL's constituents and an advisor of NCSC-NL. NCSC-NL constituents are central governmental and vital infrastructure organisations. The results from the survey and interviews suggest that such a gap exist. To help NCSC-NL in bridging this gap a model is presented. This model is based on the best practices found in the theoretical background and the findings from the advisor interview.

Table of contents

1. Introduction

Generative Artificial Intelligence (generative AI) is a type of Machine Learning (ML) that is trained to generate new data based on a combination of training data and user input [1]. This new data is often text, images, video or audio. It is a technology that is changing rapidly [2] and has the potential for misuse on a large scale [1]. Houde et al. [1] provide an example of both of these aspects. Houde et al. [1] describe a scenario (Scenario 3.2.2) where an insurance claim is denied because of a falsified video of the claiming party smoking. This scenario is somewhat similar to an attack that actually happened in 2024 where a finance worker paid $25 million to malicious actors after multiple colleagues were deepfaked in a meeting [3]. So, a scenario thought up by experts in 2020 (Scenario 3.2.2 [1]) is quite similar to an actual case in 2024, that underlines how fast the technology changes and improves.

Another way generative AI could have a cyber security impact is by aiding in programming. Phung et al. [4] state that *"results show that GPT-4 drastically outperforms ChatGPT (based on GPT-3.5) and comes close to human tutors' performance for several scenarios."* [4, p. 1]. The usage of such models in programming can lead to unsafe code being created. An example of this can be found in the work of Yang et al. [5]. In that paper they present a method of stealthy backdoor injection. A backdoor is a piece of code that provides an attacker access to the system that code is running on, a stealthy backdoor is a backdoor that is not detected by code analysis tools.

Fang et al. [6] have recently gotten Open AI's ChatGPT to autonomously hack a website. In their paper they also mention that the free online ChatGPT version, struggles much more with this. Underlining the speed of change, as these models were released less than a year from each other as can be seen when looking at the release information of ChatGPT [7] and the newer version [8]. The paper by Fang et al. also underlines that everyone who has a little bit of a budget can now use automated tools and get results that previously were only attainable with very large budgets or a lot of knowledge.

These examples show that generative AI can have an impact on cyber security through rapidly providing new capabilities to malicious actors.

The National Cyber Security Centre (NCSC-NL) is a governmental organisation in The Netherlands which has the mandate of aiding the constituents in increasing their cyber resilience. The constituents of NCSC-NL are governmental organisations and organisations

that are part of the Dutch critical infrastructure. This mandate originates from the *"Wet beveiliging netwerk- en Informatiesystemen"* (Security of Network and Information Systems Act)(WBNI), which is the Dutch law implementing the EU NIS directive [9, 10, 11] .To achieve this mandate, NCSC-NL informs its constituents through three methods: *i)* advisories about exploits and vulnerabilities and the way they are abused *ii)* publications, and *iii)* custom advice on specific cases. Advisories are detailed write-ups about vulnerabilities that have been recently discovered. In these advisories they provide the chance and damage of the vulnerability. They also explain what exactly is the problem as well as with what product and versions. Publications are about a wide range of topics, all related to cyber security, and all aim to increase cyber security resilience.

Cyber resilience is "*the ability to anticipate, withstand, recover from, and adapt to adverse conditions, stresses, attacks, or compromises on systems that use or are enabled by cyber resources*." [12]. For NCSC-NL to perform its mandate of increasing cyber resilience, it needs to know what methods can be used to cause the adverse conditions, stresses, attacks or compromises on cyber resources. One of the methods NCSC-NL has identified in being able to change the cyber security landscape is generative AI. NCSC-NL has noticed the rapid change of these technologies' capabilities and wants to inform its constituents well on this topic. It is unclear to NCSC-NL if the constituents are aware generative AI and the challenges it might bring. NCSC-NL has not published anything on generative AI yet but it might have to as NCSC-NL itself does see the scientific papers, which show that this technology might change the cyber security landscape [2, 5, 6, 13].

The main point of contention NCSC-NL deals with is that while scientific papers show new capabilities and a rapid rate of change, there are questions about if these capabilities and technologies are actually being used in practice [2, 5, 6, 13]. This however also shows a problem the NCSC-NL is facing as when these technologies and capabilities are actively used and misuse can be corroborated by the constituents, it might be too late for NCSC-NL to inform its constituents. Ideally NCSC-NL is ahead of the constituents on emerging or rapidly evolving technologies, so it can do its mandate of increasing cyber resilience well. In cases where the technology is emerging or rapidly evolving, the scientific papers might be a glimpse into the future of what might happen.

The goal of this research is to help NCSC-NL to do its mandated task of informing its constituents well, specifically on the topic of generative AI. To achieve this goal first we will

look at where the constituents are at when it comes to cyber security and generative AI. This will provide an indication for NCSC-NL to determine what steps they think need to be made to increase their cyber resilience. Besides this, we will ask what the constituents currently think of the information being provided by NCSC-NL. This opinion will be split in two parts, the content of the publications and the information in the publications. We will also look into whether scientific papers can be a good source of information for NCSC-NL, when it comes to anticipation in the case of emerging technologies. Information from papers will also be compared with the information seen in practice by the constituents of NCSC-NL. This is done to see if there is a gap between the information that can be seen in research and what is happening in practice. According to Tkachenko et al. [14] such a gap could be called a research practitioners gap. To summarise, the goal is to gain insight in how NCSC-NL can spread generative AI information the best. This will be done by looking into the challenges NCSC-NL faces in spreading this information. Thus, this brings us to the main research question.

"How can the challenges of NCSC-NL, regarding their statutory task of to increase digital resilience be characterised, when looking at the case of generative AI?"

This research question will be answered by answering the following research sub-questions:

1. How do constituents perceive the reference publication [15] in terms of i) clarity and ii) usefulness?
2. How accurately can constituents determine their cyber resilience with regards to generative AI?
3. What type of information with regards to generative AI are the constituents missing or lacking and how will this information improve the cyber resilience of the constituents?
4. What are the common denominators with regards to the missing information from the NCSC-NL within or across sectors or types of organisations?
5. What adaptations need to be made to make the best-practices from other disciplines work at NCSC-NL?

These research questions are formulated with the following two assumptions in mind:

1. There is a gap between the publications by the NCSC-NL and the information need (perceived or yet unaware) of its constituents with regards to generative AI.

2. It is possible to shrink the gap between the NCSC-NL's publications on generative AI and the information needed by its constituents to increase their cybersecurity resilience.

These assumptions are based on the opinion of a researcher and an advisor from NCSC-NL. These people are well placed within the organisation to determine such assumptions as one of them sees the scientific side, and the other is more in contact with the constituents. The combination of these different perspectives and opinions lead to the assumptions. The assumptions were one of the main driving factors in doing research on this topic.

We answer the research questions by first discussing the scientific building blocks in the Section 2 in which we try to make our working definition of the research-practice gap as well as show the research-practice gap in the real world. We will also look for best-practices for shrinking this gap from the literature. Afterwards related work is discussed. In the Section 3 the research questions are linked to the way we will answer these questions. Afterwards, in the Section 4, the results from the survey and interviews will be presented and tied back into how they answer our research questions. Then a model is be presented in Section 5 which aims at aiding NCSC-NL in shrinking the research-practice gap. Lastly, limitations will be discussed in section 5 and a conclusion will be drawn in the Section 7.

2. Theoretical background

At the starting point of this thesis, the assumption was that there was a gap between the knowledge of the advisors of the NCSC-NL and the knowledge of the constituents. To make sure we used the right type of knowledge gap we dove further into the types of gaps. We found three types of gaps in the literature that might be applicable here, a knowledge gap, a research gap and the research practitioners gap. After careful consideration between these types of gaps it became clear that the research-practice gap is the one that applies best. Knowledge gaps focus more on internal company knowledge and how to spread that through knowledge management systems [16] and a research gap is *"When the ability of the systematic reviewer to draw conclusions is limited*." [17, p. 1]. The research-practice looked the difference between the knowledge of scientific literature, and what was actually being used and known by practitioners. This seemed to characterise the gap the NCSC-NL saw the best.

A challenge in choosing the research-practice gap was that it was hard to find a definition of what exactly this gap entails. All definitions are a description of the problem or situation they are examining, hence they are more of a context description than a definition. The closest to a proper definition that was found was *"It does not take much effort to picture discussions of academic–practitioner relationships that focus on the "gap" (or divide or similar metaphors) between academics and practitioners, between rigor and relevance, between theory and practice or similar terms"* [18, p. 2]. The main point of this definition is that there are two distinct groups, practitioners and academics and that there is a disconnect between these two groups. However it does not clearly note what this gap is apart from two small examples and does not provide a clear picture of what exactly this gap is.

The reasons for a lack of a clear definition is likely two-fold. Firstly, not everyone uses the exact same terms, as can be read in Tkachenko et al. :*"While conducting the review, we also noted that scholars often employed the terms gap, divide, or gulf (or similar metaphors) interchangeably, that is, implying the same meaning. Similarly, the verbs closing, narrowing, and bridging were employed without explicating the difference between these actions"* [14, p. 9].

Secondly, the definition is contained in the term itself; it is a gap existing between the researchers and practitioners. In Tkachenko et al. they also mention that *"While reviewing the literature, (...) we came to an understanding of the absence of a comprehensive model that*

would holistically portray the key components affecting interplay between research and practice" [14, p. 7]. To help with this issue, they provide a model This model is a largely visual model, see Figure 1.

Figure 1: Tkachenko et al.'s model of their view of the research-practice gap [14]

In Figure 1 they try to visualise the interaction between research and practice. It shows them both being at opposite ends of the model with the so-called "Yin-Yang" sign in the middle. It shows the interaction between both sides through two means, the process and product of knowledge production, according to Tkachenko et al. these are the main reasons for the gap existing. The main use of this model in this thesis is to serve as an example of what a model in the research-practice gap can look like and what aspects can be a cause of the gap.

The fact that there is this lack of a pure definition until 2017 when Tkachenko et al. published is something interesting to note. This is because even in one of the first papers namely Starkey and Madan [19] in 2001 they cite sources from the beginning of the 2000s or the end of the 90s. This means that for almost 20 years, there have been dedicated journals for the contributions to understanding the research-practice gap, yet none of these give a clear and concise definition [20]. With none of these papers deeming it necessary to give their exact definition, nor quoting any definition it seems very likely that the reason for a lack of a clear

definition is that the definition is encased in the text of the term. The term describes exactly what it is, even when using synonyms.

The working definition of the research-practice gap for this thesis is as follows:

> *The disconnect between information researchers publish and what practitioners can use.*

It encompasses all terms of the research-practice gap while providing more information about it having to do with information being published and that it is information the practitioner can later use. The main point of this section of the theoretical background was to clarify what the research-practice gap is and to provide the definition for it for this paper.

2.1 Examples and best-practices of research-practice gaps

While the paper by Starkey and Madan [19] is one of the first to name the Research practice gap, they mostly focus on just one aspect of it; the alignment of stakeholders. The later papers [20, 21, 18, 22, 23] do agree with this first one on the importance of stakeholder alignment, but as mentioned in Tkachenko et al., they take a more holistic perspective when making their model [14].

Bansal et al. [20] start their paper off by summarizing a few ways in which the research-practice gap has arisen, they name that researcher prefer to produce new knowledge rather than spreading it or translating it [23], researchers are incentivised to produce research [24] instead of talking with practitioners, researchers and practitioners convey information in different manners and they use different language and strategy [22, 25] and lastly researchers and practitioners hold different perspectives on the nature and acquisition of knowledge [26]. Lawler and Benson [21] add to these points by saying that one of the reasons they see for this gap existing is that academic literature results can be hard to understand, not particularly applicable or conflicting. They actually found that *"a large number of practicing HR* [Human Resources] *professionals either do not know the research findings and actually believe the opposite."* [21, p. 2], this result was confirmed again in 2015 when KPMG did a survey on European executives [27]. Lawler and Benson also found in their research that the side being blamed for this gap existing is mostly the researchers. They note that this can partly be attributed to the communication style, however it is also partly down to the subject of the studies being done. Some of the research being done has no real world implications or applications. They note that while some people say the gap is widening, they question if research ever was very close to practise. Bartunek et al. [18] agree with some of the points

already mentioned, but add that time dimensions also should not be forgotten, the timelines for researchers and practitioners simply is not the same. While researchers often look far ahead, practitioners look more at quicker wins and thus often less far into the future.

Besides mentioning causes for the research-practice gap, these papers also note steps they have taken to try and shrink it. In the case of Bansal et al. they founded an organisation the Network for Business Sustainability (NBS) [20]. The reason for founding this organisation is to make it easier for both researchers and practitioners to exchange their knowledge. One of the ways they try to shrink the gap is by taking one or more papers, and making an easy-to-grasp model or framework out of them. By making these models and frameworks they deem that the research is easier to apply for the practitioners as the practitioners do not need to analyse the paper and decide how to implement it first.

Lawler and Benson [21] write about an organisation of a similar nature, it is called the Centre for Effective Organizations (CEO) at the University of Southern California. It is an organisation that aims to align research and practice and has done so for the past forty years. This organisation is similar to the one named in Bansal et al. [20] in the fact that it tries to align research and practice, however CEO does it by bringing them together instead of translating research to appeal/apply to practice. They bring them together by getting both sides in a meeting and having discussions about the research being done, or the research to be done to align the perspectives of both the researchers and the practitioners. They also noted that the papers which were the closest to practice, were always written by people who were in both worlds. The most important piece of information from this paper is the way CEO works, they mention that *"Practitioners come to CEO looking for models to understand changes in practice, research skills to test them, and a window into what other companies are doing."* [21, p. 4]. This shows that while CEO is similar in the way that it tries to align research and practice, the way they do it is by helping practitioners understand research on request instead of doing it proactively.

The main point of these organisations is bringing the two worlds together, this is also a prevailing topic in other sources. Mcintyre suggests in [28] that the first step that should be taken in this is having a dialogue between the researchers and practitioners. The relevance of this idea is also underlined by the papers about the CEO and NBS as well as [19, 22, 23, 29, 30].

In addition to the broader idea of having a dialog, Sutherland et al. provide five categories of best practices in their paper [31]. They also specifically mention that *"These methods are transferable to a wide range of policy or research areas within and beyond the conservation sciences."* [31, p. 1]. The five main categories they propose are *i)* defining the project , *ii)* organising the participants, *iii)* soliciting and managing questions or issues *iv)* voting systems of what to pursue and *v)* disseminating results. The main methods that are usable in our context are defining the project, organising the participants and disseminating results. Under these main categories the most applicable principles they provide are:

- Vision: A clear understanding of the vision makes it easier to choose the method, identify potential participants and already have a desired output in mind. Sitting down with the stakeholders to make sure the vision is aligned on this project will ensure all parties know exactly what to expect.

- Scope: Scoping the project to know exactly what questions will, and will not be answered, determining this will help know what to expect of the end result.

- Number of priorities: Knowing the number of priorities will help to know roughly how large the research is, in addition to knowing how much time will need to be spent on it, which will make planning easier.

- Organising team: Knowing exactly who needs to be in a team will ensure that all necessary tasks are being done, and help to have a successful study.

- Composition of participants: Making sure the participants of the study are diverse in all manners ensures a broad knowledge base.

- Number of participants: Choosing the right number of participants is important, as when the amount increases, there is a broader base. However, this also has the downside of providing more questions and is relatively ineffective at generating output.

- Facilitation of breakout sessions: Have someone who knows a lot about the topic to help answer questions and choose the right questions to answer in workshop sessions. This helps ensure the workshops are smooth.

- Transparency and democracy: Ensuring openness and transparency amongst participants helps give the research more legitimacy. It also helps spark questions and discussions as

someone might see something slightly differently, and thus a scope adjustment might be necessary.

- Participant ownership: Ensuring the participants feel ownership of the research will make them more likely to help and also more invested in the project besides just receiving the results. It can also help enhance legitimacy and credibility even more.

Chen et al. [32] give three bottlenecks that exist in their process to keep the Institute of Industrial Systems Engineers (IISE) relevant and valuable. The three bottlenecks they see are *i)* verifying the performance improvement, *ii)* building trust with practitioners and *iii)* balancing model accuracy and simplicity. As said before they do not provide many best-practices however they do provide potential research opportunities and give case studies. Lastly the one best-practice they do provide is using data-driven decision methods to deal with the bottlenecks they are struggling with.

The current methods of the NCSC-NL do not yet involve all of the best-practices named above. The NCSC-NL's methods however do look quite like the CEO as mentioned in Lawler and Benson [21]. It gives the practitioners the information they need to be more cyber resilient, both proactively and on request like the CEO does. Besides this a bottleneck the NCSC-NL also faces is the trust of practitioners. Improving the cyber resilience of the constituents is only possible if they accept the advice from the NCSC-NL. With the NCSC-NL having its own scientific research team, the papers they write might be close to practice as Lawler and Benson mention that the closest papers were written by people who are in both worlds.

The main takeaways from this section of the theoretical background is *i)* the position of NCSC-NL is not unique and other organisations find themselves in similar situations and *ii)* previous research has presented a wealth of best-practices to address research-practice gaps. Some of the best-practices from this section will come back in the model we design to help NCSC-NL bridge the research-practice gap.

2.2 DIKW pyramid

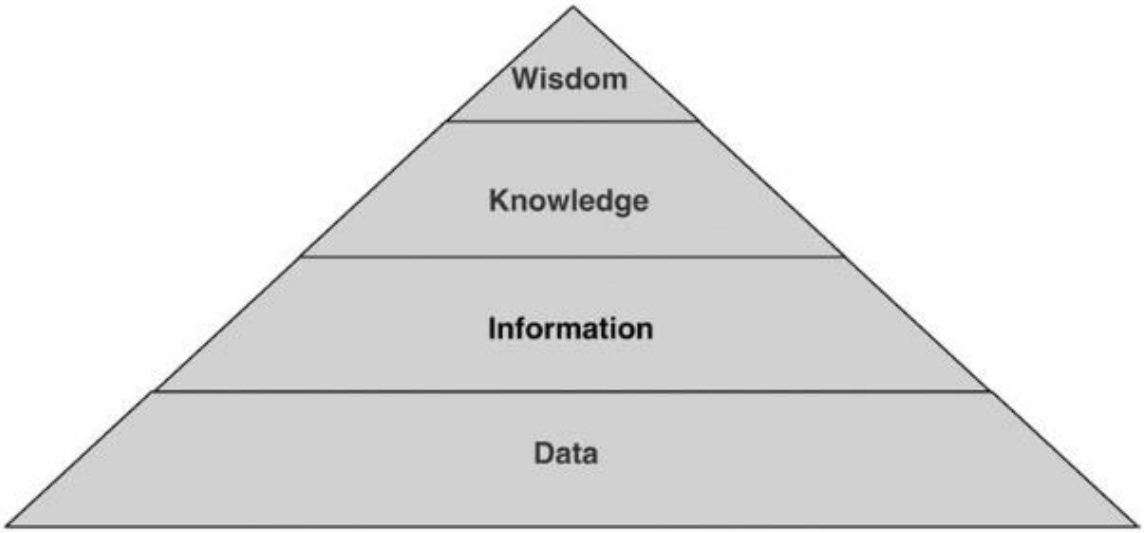

Figure 2:The DIKW pyramid

The DIKW pyramid, also known as the Knowledge Pyramid, is a model in the information and knowledge field. It is discussed and explained well by Rowley in [33] and can be seen in . It shows the relationships between data information knowledge and wisdom and that Data is the foundation of Wisdom, every step upwards adds context and the results become more valuable. The DIKW pyramid can be used to provide an abstraction of the type of information. It will be used in this thesis to provide an abstract level assessment of the type of information needed by the constituents.

2.3 Related work

Related work is limited on this topic as the research is quite narrow. To avoid repeat work, at the beginning of the process NCSC-NL reached out to international partners to see if they have done any similar research. All of the partners relayed the information that they have not done any similar research.

The work of Hare [34] is probably the most closely related work. They suggest that public private partnerships can be an opportunity in the cyber security space. This research was conducted at the Department for Homeland Security in the USA. The main takeaway is that the Department for Homeland Security could improve collaboration by providing important cyber security information to the private sector. This is what NCSC-NL is trying to improve with this research.

Carr [35] mentions that the private sector does not like to accept responsibility for national cyber security and that acknowledging that is an important step in increasing cyber resilience in these organisations. They explain that putting aspects that are important to these private organisations first helps with getting information across.

Stoddart [36] outlines how the UK National Cyber Security Centre (NCSC-UK) works. It outlines the constituents, which is critical infrastructure as well, and how NCSC-UK helps them increase their cyber resilience. An important difference between NCSC-NL and NCSC-UK however is that NCSC-UK is a part of Government Communications Headquarters (GCHQ), a UK intelligence agency whereas NCSC-NL is a separate entity from the Dutch intelligence community.

As can be seen from these related works, this research is in an area in which there has not been much research. All related works are only a little related but are not on the exact topic of this research.

3. Methodology

Information provision is one of NCSC-NL's core tasks. NCSC-NL is currently seeing the changes in the cyber security landscape when it comes to generative AI technologies and its capabilities. NCSC-NL is also wondering if it is serving its constituents well through the information it is currently providing when it comes to generative AI. To aid NCSC-NL in answering the main research question of how to characterise the challenges of NCSC-NL when it comes to generative AI the following steps need to be taken: firstly, a literature analysis already is done in the previous section to see exactly what type of gap this situation is. This is done to gain understanding in this gap and to see if there are any best practices in bridging this gap. This analysis also provides a foundation on which the definition was based. The analysis also showed that there is a research-practice gap and that this accurately describes the gap NCSC-NL sees. It also provided knowledge about the best-practices that are currently out there to help bridge a research-practice gap.

NCSC-NL commissioned this research. This inevitably leads to some bias. This bias was limited as much as possible through multiple means. Firstly the research was conducted within the scientific research team, they helped to maintain scientific rigour by helping keep the methodologies robust as well as keep the study ethical. Besides this, the team is very familiar with scientific methods and thus knows that the findings can include anything, but also nothing. To conduct the research good methods of scientific research were used, an example of this is for the survey first the goals were selected [37]. These goals then lead to the questions that needed to be answered. NCSC-NL did not provide a preferred result of this research as to not interfere with the research process.

3.1 Survey

To answer the research questions a multi-step approach is chosen. First, a survey is conducted amongst the constituents of NCSC-NL. After an in-depth interview is conducted with some of the respondents of the survey. Lastly a second interview is conducted to gain more understanding of NCSC-NL's standpoint and processes. The combination of these steps is chosen to limit the downsides of each method, as well as gain the most understanding possible [37]. The data collected is a combination of quantitative and qualitative data. The data is gathered through a survey and two types of interviews. The survey mostly collected quantitative data with a couple of open ended questions for the possibility for qualitative data.

To understand the trends seen in the data from the survey qualitative data was collected from two types of interviews. This combination of using a combination of data is often used and helps to see and understand trends [38].

Research Question	Answered through
1. How do constituents perceive the reference publication in terms of i) clarity and ii) usefulness?	Survey
2. How accurately can constituents determine their cyber resilience with regards to generative AI?	Survey
3. What type of information with regards to generative AI are the constituents missing or lacking and how will this information improve the cyber resilience of the constituents?	Survey In-depth interviews
4. What are the common denominators with regards to the missing information from the NCSC-NL within or across sectors or types of organisations?	Survey In-depth interviews
5. What adaptations need to be made to make the best-practices from other disciplines work at NCSC-NL?	Theoretical background Advisor interview

Table 1: Research questions and the methods for answering them

The survey method is chosen as it is method for studying the knowledge and attitude of the population towards the subject [39]. Both are important factors for research questions 1, 2 and 3, why. The knowledge is also important in research question 4, why. To answer research question 1 a reference publication is attached. The reference publication can be seen in Appendix D. This publication was supposed to be an NCSC-NL publication but due to time considerations the National Security Agency (NSA) / The Artificial Intelligence Security Center (AISC) publication [15] is used. The publication will from here on be named the reference publication. The reference publication is chosen as it is technical publication on the topic of AI and was published recently. Unfortunately the topic discussed in the reference publication is not generative AI so it is a little removed from the topic of this research. The NCSC-NL publication was also supposed to be a technical publication which is why the reference publication is used. The reference publication is used as it is created by a governmental agency like NCSC-NL is. It also includes a course of action to take to limit problems, which is something NCSC-NL also includes in its publications. These reasons make the publication as close to a NCSC-NL publication without being one.

The targeted respondents for the survey are all constituents of NCSC-NL. The survey is sent out using the NCSC-NL research emailing list, which covers roughly 280 email addresses from 120 constituents. These people have voluntarily signed up to receive emails from the research team within NCSC-NL. Besides this the survey was announced at a meeting of (vice)-chairpersons of Information Sharing and Analysis Centers (ISAC). These ISACs were later contacted by a relation manager from NCSC-NL. This involved roughly 225 CISO's of constituents. There was some overlap from both methods of contact, which leads to a rough estimate of 280 constituents being contacted and 380 individual emails being contacted. In total there were 30 responses, this provides a response rate of 8%. The survey was open for a month and people from the NCSC-NL research emailing list received a reminder email after two weeks. The people that left their email in the survey results did not receive a reminder. The survey took 5-7 minutes to fill in.

3.2 In-depth interview

To get more background information an in-depth interview with some of the respondents is conducted. Wilson [40] suggests that semi structured interviews are good when you want to give interviewees the option to raise new issues, get a deeper understanding in the answers and clarify answers. These options are important to this research to get a good understanding of what information is missing. A semi structured approach was therefore chosen. The interviews are not labelled but instead thematic analysis is used. Kiger and Varpio [41] suggest that thematic analysis is designed to search for shared or common meanings but less effective for finding individual experiences. The goal of these in-depth interviews is to gather what constituents are missing, that is a shared meaning. The interviewees were selected from a list of survey respondents. All of the interviewees marked that they were open to be interviewed in the survey. There was no further selection made as only five people were open for an interview and left contact information. Three of the interviews were conducted online via Zoom, two were in-person as the respondents were in the same building. All interviews were conducted between June 3rd and the 25th of June. The interviews were an hour. Of this hour 45 minutes were allocated ahead of time for all the main questions. The questions can be found in Appendix B, together with the time allocation.

To deal with ethical considerations, only respondents of the survey that answered they would like to be interviewed were contacted. They are given 14 timeslots of multiple hours and were asked to pick the two that suited them best. They were not sent a reminder email if they did

not respond to the first email. The interviews were recorded with the consent of the interviewees, these recordings were used to make a transcript, after this transcript was made the recording was deleted. When this research is completed, the transcripts will be deleted.

3.3 Advisor interview

The second interview type is an interview with an advisor of NCSC-NL. This person is subject matter expert on cyber security and the methods of getting to a publication. In the interview the advisor is asked questions about where they think the constituents are at with regards to generative AI. Besides this they are asked about the usage of scientific papers in the publication process and what they think of the research-practice gap. Lastly some findings are presented to the advisor, this is done when all the answers are already given to not influence further answers. Sharing the findings is done to inform the advisor on the standpoints of the constituents and to reflect on the answers the advisor gave in the first part. The main focus is on the publication process and the research-practice gap. The interview is analysed by looking at the answers the advisor gave which were confirmed by them. The interview mostly serves as input for research question 5 and the model. The goal of this interview is to be able to answer research question 5 by gaining insight in the process of making publications to see where best-practices from literature can be applied. This interview was conducted last as to not bias the rest of the research approach. All other interviews were already conducted and all data from the survey was already collected.

The interviewee agreed to be interviewed and to the interview setup. The interview data will be deleted on the completion of this research.

3.4 Model

Using the data collected by the survey and interviews a model is designed. This model is aimed at bridging the research-practice gap as it is defined in this thesis. The survey and advisor interview combination are used to identify the gap and get an idea of its size. The survey and in-depth interview combination are used to identify the gap, as well as to see on what topics this gap is. The information of the topics is used in the process of designing the model to see if these topics would come to light if the model is used in the future. The model is designed by looking at the best-practices from the literature in addition to the information from the advisor interview. The main focus of the model is to look at what best-practices can be added to the process to address parts of the gap. Besides this an already existing model,

signposts of change [42], is used to address the rate of change and uncertainty of generative AI. Since it is unknown if all research will be used in practice later and the rate of change is high, a model that splits up aspects is used. Signposts of change are a useful tool to monitor and evaluate changes and can help depersonalise arguments by looking at what is provable [42]. Signposts of change also help list observable events that you expect to see in a situation. Besides this, in this situation signposts of change help NCSC-NL in to deliver the information in a timely manner. By splitting it, it allows NCSC-NL to deliver the information before a lot of the constituents have had to deal with a certain capability. It also helps NCSC-NL in determining what capabilities will actually be seen in practice and what remains theoretical. By splitting up the capability in smaller sections, NCSC-NL can see what capabilities are progressing and the constituents thus need to be informed on. The model was planned to be validated by a third party expert due to time constraints this did not happen. This third party is an advisor from an organisation similar to NCSC-NL who also has a PhD in AI and is well known with scientific papers on this topic. The combination of these factors made this person very well suited to verify the model's feasibility.

After this methodology section the results will be presented in the Section 4. The results will be presented in the order the methods were presented here. After the results section, the model will be presented in Section 5. This model is based on the gap found from the survey and interviews as well as the best-practices found in the literature. In Section 6 the research questions are answered by analysing the results, limitations are presented and throughout some future work will be presented.

3.5 Data collection

3.5.1 Survey

In the survey we first ask general questions about the organisation for which the survey participant works. These questions range from the name of the organisation, to the amount of IT security personnel to what sector the constituent is active in. This section is to help answer research question 4 as it provides the possible common denominators. The full survey can be seen an Appendix A.

In section 2 the survey participant is asked to answer some questions about the reference publication [15]. This publication is interesting as it is a technical publication from

organisations similar to the NCSC-NL and addresses a topic close to generative AI. The reference publication is written by the AISC, NSA and NCSC-UK and others to inform their constituents and the public on the topic of safe deployment of AI systems. The publications provides insight and measures on this topic making it interesting to the NCSC-NL constituents as well. The participants are first asked to rate this publication on whether or not it was actionable, useful, clear, informative and complete. They are asked to rate it on a scale from 1-5 with 1 being they disagree with the statement and 5 being they wholly agree. Lastly they are asked to give it a rating from on 1 through 5 stars. An open field for any notes was also given. These criteria are chosen as we deemed these factors together would lead to the rating of the reference publication. This section helps answer research question 1 by providing the perception of the publication.

In section 3 the participants are asked more questions about the reference publication, in addition to this they are also asked questions about what they usually do with NCSC-NL publications. The reference publication states in its introduction that it has three goals. The participants were then asked if they think these goals are achieved. Afterwards there is an open field where they are asked if they think this publication achieved its overall goal. Lastly the participants are asked on what type of organisations, what sector and what roles within the organisations the publications is targeted and how they usually use NCSC-NL publications. This section helps answer research question 1 by seeing if the respondents think the goals are met. The section also helps answer research question 3 by providing a base of what the constituents are missing information on, through the open ended questions.

In section 4 the participants have to rate their own knowledge on nine different subjects. These subjects are all related to generative AI. The reason for asking these subjects is to accurately gauge the actual cyber resilience and knowledge with regards to generative AI. This is part one of two of answering one of the research questions. The participants were given the options *"Unknown, I looked it up once, The knowledge is present in the organisation, I could apply it and I am an expert at it"*. These options are chosen as they were deemed as easier to answer than rate your knowledge on a scale from 1-5 by giving them something which they can imagine easier. The nine technologies the participants were asked about are *"prompt injection, automated hacking, deepfakes, LLM*[Large Language Models] *finetuning, model stealing, vector databases, Mistral 22b, LLaMa 2 and Gemini".* Large Language Models (LLMS) was named by their abbreviation. The listed technologies are seen as novel technical parts of generative AI [1]. The technologies chosen have also gone

through a round of feedback by three people within the NCSC-NL. This section helps answer research question 2 by providing a measure of how well the constituents know generative AI technologies. This knowledge is later compared to the knowledge of the next section where the respondents provide a grade for their own cyber security and risk management. This section also helps answer research question 3 by providing a base of what generative AI technologies the constituents are missing information on.

In the final section the participants are asked about risk management within their organisation and the role generative AI plays in it. They first are asked however to rate their maturity on a scale from 1-5 with regards to risk management and cyber security. The reason for these questions was twofold. Firstly it gives us another point to segment the groups on and secondly it is to see if their assessment lines up with ours with regards to generative AI. After these questions, the focus of the next questions is on the use of generative AI within their organisation. Firstly they were asked to tell how many generative AI applications were used within their organisation, 0,1-5,5-25,25-50 or 50+. The follow-up questions are, to your knowledge, which are they, what departments use them and lastly what limitations have been put on the use of generative AI applications. These questions mostly serve as a measure to see how widely used these systems are and what is being done with regards to limitations on them. This section helps answer research question 2 by comparing it to the results of the previous section.

3.5.2 In-depth interview

The structure of the interviews with constituents is determined by first selecting the three main pillars of the interview. These pillars are: *i)* generative AI within the interviewee's organisation, *ii)* the role of the NCSC-NL and *iii)* the interviewee's opinion on generative AI. The idea behind these pillars is twofold, it is both to verify some of the findings from the survey as well as to collect more data. The interview guide for the in-depth interview can be found in appendix B

In the first pillar, generative AI within their organisation, there are 9 main questions, of which 4 are deemed to be the most important and will definitely be asked every interviews. On top of the 9 main questions, there are 4 clarifying questions to ensure that all points of interest are covered. These questions are only asked if the answer to the main question does not yet include the answer to the clarifying questions. This pillar is chosen the gain a better understanding about the use of AI within the organisation. That will help with better

understanding the position of generative AI in the constituents. After deliberation with some of the advisors within the NCSC-NL this will help us measure the gap. This section aids in answering research question 3 and 4. The section helps answer research question 3 by providing insight in what the constituent already uses, as well as what is already known. The section helps answer research question 4 by providing a base of what information is missing and to see if there is any relation between what is missing and the constituents being interviewed.

In the second pillar, the role of the NCSC-NL, there are 5 main questions, of which 3 are the most important and there are another 4 clarifying questions. While there are fewer questions in this pillar, it is the pillar which is allotted the most time. This is because the data collected, the possible methods of shrinking the gap, is new and important in this research. This section helps with answering research questions 3 and 4. In this section the constituents are asked what they want from NCSC-NL with regards to generative AI information. This helps with answering research question 3 as it shows what information is missing. The information this section adds to help answer research question 4 is if there is any connection between the missing information and the constituents being interviewed.

In the final pillar, the interviewee's opinion of generative AI, there are only 4 questions, 2 of which are most important and only 2 clarifying questions. This pillar is mostly chosen to get a feeling for what will happen in the future specifically with generative AI in the cyber security sector. This will help paint a picture of what the practitioners think the future of generative AI looks like, this can then again be compared to the opinion of the advisors to determine the gap between the two standpoints. This section does not help answer research questions but is still important. It is aimed at gathering this information so NCSC-NL know what the opinion of the practitioners is when it comes to generative AI.

4. Results

The results from the survey are based on 30 respondents from 28 organisations. It is also good to keep in mind the spread of both the sectors and roles as seen in Figure 3 and Figure 4.

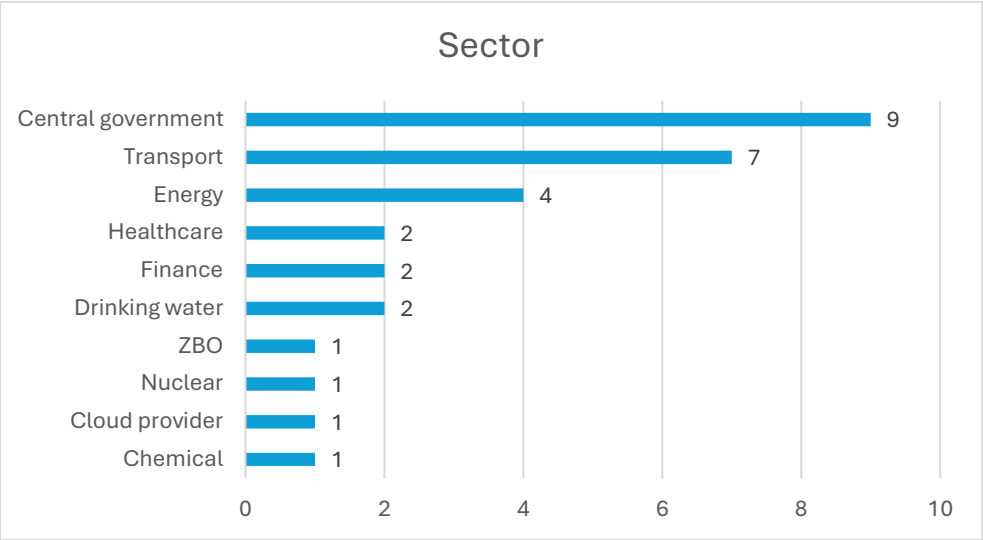

Figure 3: Distribution of sectors

As is can be seen in Figure 3 the distribution of sectors leans a bit on two sectors: central government and transport. Together they make up just over 50% of respondents.

Figure 4: Distribution of roles

In Figure 4 it can be seen that 8 CISO's have responded to the survey. This number shows that reaching out to this group specifically through the ISAC's has had the intended effect. The distribution of both the sectors and roles will later be discussed in Section 6, however it does provide good context for the following sections of the results.

The spread observed on both the roles and the sectors means the respondents are not representative of all constituents. The findings and conclusions are therefore made for the respondents and not necessarily for all constituents of the NCSC-NL.

As mentioned in the Section 3, the first part of the survey aims at getting background information of the organisations. The number of employees as well as revenue/budget and IT-budget mostly serve as segmentation information. The only thing that arose from the combination of revenue/budget and IT-budget is that the IT-budget seems to roughly be 10% of the total budget of an organisation. From all the respondents, 73% have a CIO, 83% have a CISO and 77% have a SOC. These points paint a picture of the average survey respondent.

In the next section information about the reference publication is gathered [15]. Among the datapoints collected was a grading of the publication on a scale from 1-5.

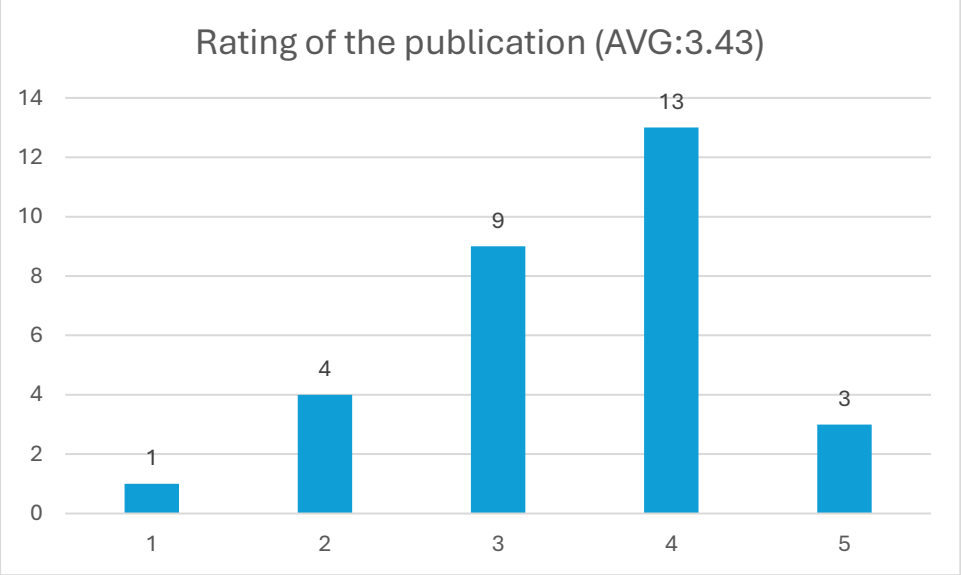

Figure 5: Rating of publication

As shown in Figure 5, the average rating of the publication is 3.43 with the mode being 4. The averages of the rating of certain aspects of the publication were roughly around the rating of the publication. The averages of the individual parts were 3.20 for completeness, 3.46 for usefulness, 3.83 for clarity, 3.80 for being actionable, 4.07 for the technical level and 4.03 for

how informative the publication is. The fact that all of these individual aspects of the publication are close to the rating of the publication suggests that the aspects chosen to represent the publication overall are the right ones. This is corroborated by the fact that the distribution is very similar as well. All results from the aspects look very much like a normal curve.

Aspect	Completeness	Usefulness	Actionability	Clarity	Informative	Technical level
Average score	3.20	3.46	3.80	3.83	4.03	4.07

Table 2: Average score of aspects from low to high, left to right

In the publication they state that "*The goals of the AISC and the report are to: 1. Improve the confidentiality, integrity, and availability of AI systems; 2. Assure that known cybersecurity vulnerabilities in AI systems are appropriately mitigated; and 3. Provide methodologies and controls to protect, detect, and respond to malicious activity against AI systems and related data and services.*" [15, p. 1]. The respondents found goal 1 least present with 5 respondents giving it a 1 and the main feedback in the open field regarding these questions mentioning that the CIA triad wasn't mentioned at all. Goal 1 is about the CIA triad. The CIA triad stands for Confidentiality, Integrity and Availability [43]. These are terms are considered the heart of information security. These parts are often looked at when looking at incidents to see what was impacted. The averages however were again roughly in line with the average of the publication with 3.17 on the CIA triad, 3.46 on the known vulnerability mitigation and 3.37 on providing methodologies.

Overall the ratings suggest that the publication is well received and that it largely meets its goals. The open ended questions provide an interesting caveat however. One of the main points of feedback within the open ended questions was that the publication, while achieving its goal, does not focus enough on AI. With the publication being titled *"Deploying AI Systems Securely"* this is quite unexpected feedback. Besides this being named, the main point of feedback is that it is quite generic. A peculiar finding from the open ended questions is that some people note that it is not interesting, with others mention that it provides new insights and think the information is valuable. While these answers are contradictory, the publication does mention that *"These best practices (...) are not applicable for organizations who are not deploying AI systems themselves and instead are leveraging AI systems deployed by others."* [15, p. 2]. The main point added to the note that it is not interesting is that they

are either not that technologically advanced enough or that they use commercially available AI systems. This lines up with the quote from the reference publication and means that maybe those respondents were not the target audience.

The respondents were also asked about certain generative AI technologies. They had to rank these on the scale *"Unknown, I looked it up once, The knowledge is present in the organisation, I could apply it and I am an expert at it"*. The technologies were prompt injection, automated hacking, deepfakes, LLM finetuning, model stealing, vector databases, LLaMa2, Mistral AI and Gemini.

		First	Second	Third	
	Unknown	I looked it up once	The knowledge is present in the organisation	I could apply it	I am an expert at it
Prompt Injection	3	5	16	5	1
Automated Hacking	2	6	12	10	0
Deepfakes	1	7	13	9	0
LLM Finetuning	7	7	10	6	0
Model Stealing	10	7	9	4	0
Vector databases	13	5	7	5	0
LLaMa2	15	8	5	2	0
Mistral AI	16	8	4	2	0
Gemini	10	8	7	5	0
	29%	23%	31%	18%	0%

Figure 6: Colour coded table of the technologies

In Figure 6, the table with colour coordination of the technologies is shown. The cells with the blue background have the most answers for a certain technology, the orange cell is the second most, and the green is the third most, other cells are left white. As clearly visible in Figure 6, we deem there to be two distinct groups, the first group is the more general techniques consisting of prompt injection, automated hacking, deepfakes and LLM finetuning. In this group the most answered option is *The knowledge is present within the organisation* and the second and third option usually being *I looked it up once*, or *I could apply it*. This group is of well-known generative AI technologies, among the respondents there was only 1 expert, in prompt injection, and very little people unknown with these techniques.

The second group, of the remaining techniques is less known according to the results. These techniques are more niche. The main answer is to these techniques is *Unknown* with the second option either being *I looked it up once* or *The knowledge is present within the organisation*. These technologies are more technical and implementation focussed .The only real outlier compared to expectation is Gemini. It also is slightly the odd one out compared to the rest of the list, it is not a technique on its own, and not an opensource model. It is Google's answer to ChatGPT and Copilot. There is also a business version for called Gemini for Google Workspace. The fact that this model and suite is unknown to a third of the respondents is interesting to see.

As mentioned in Section 3, these techniques were asked as a measure to gauge the cyber resilience of the constituents when it comes to generative AI. The results suggest that a lot of information is present within the organisations, however the difference between that option and unknown is quite small. The option for *I looked it up once* is further behind unknow. This suggests that technology is most likely to be known in the organisation, or unknown. The information being present in the organisation is definitely very good to see.

An explanation for these results might come from one of the others questions posed later on. This is the question of what generative AI techniques are used within your organisation. Where 50% of results say they use 1-5 generative AI applications, and the most mentioned one is Microsoft Copilot. Techniques like model stealing, vector databases, LLaMa2 and Mistral AI are most relevant if they make or deploy their own models.

The result of some but not very deep knowledge is an interesting change from what the respondents note about their own cyber security and risk management. The respondents gave their own risk management maturity an average of 3.38 and their cyber security maturity a 3.31. These numbers don't necessarily conflict with the results from the generative AI technologies knowledge. Given the grades, the knowledge of the technologies falls within the realm of expectation, it is however on the lower side of the range of expectation. A real conclusion cannot be drawn from the data as the results are not to the contrary of one another. It did become clear later that of the interviewees, none used custom or in-house models, they all used models from third parties. This might explain the lack of knowledge in the second grouping of techniques.

The knowledge of the constituents was expected to be a little greater than the results suggest. In short, the constituents appear to not be cyber resilient when it comes to generative AI.

When asked in the interviews, none of the interviewees suggested they were resilient with regards to generative AI. While this might sound like a very bad scenario, during the interviews participants were asked if they have seen any incidents with which generative AI is used and just the one suggested that they did notice a spike in the amount and quality of SPAM. This did lead to the question of exactly what information the constituents need to be cyber resilient when it comes to generative AI.

The fact no real attacks have been seen yet is in contrast to the capabilities shown in scientific research. These papers cover a range of topics from LLMs hacking websites on their own [6], using an LLM as a hacking assistant [13] to poisoning prompts from generative AI applications [44]. Here quite clearly a disconnect, a gap, is visible between the research and practice. The information shown in the research is not all used in the real world. There are however known uses of LLMs in practice as shown by Microsoft [45]. This provides an opportunity for NCSC-NL to inform the constituents before the capabilities are in active use.

A thing to note from the results of the question about the number of generative AI applications in use is that some respondents noted they use 25-50 or even over 50 applications that make use of generative AI. One of them notes that there are some in the Software as a Service (SaaS) products they use, the others did not clarify. While it might be possible that they have that many applications using generative AI, it seems more likely they confused generative AI products with products using AI. The thought behind this standpoint is that even during the interviews some interviewees confused Machine Learning (ML)/AI with generative AI.

The usage of generative AI within the constituents appears to be quite widespread, 73% of respondents admit to using 1 or more application with generative AI in it. From the interviews this was 4 out of 5 interviewees. The implementation and reasoning behind their use however vary. Most interviewees stated that it increased productiveness throughout their organisation. This is often touted as the reason to use generative AI. None of the companies however let the employees have free reign over the generative AI tools, all companies had policies in place on what processes can and cannot use generative AI. Mostly, this meant that processes using customer information or critical company processes did not allow for the use of generative AI tools. When asked more about this they note that data privacy and AI assurance were the main causes of this. Data privacy as they were afraid of information leaking or being send abroad and AI assurance as they did not know how the generative AI

came to its result. The interviewees were afraid of injecting wrong information into a process and some have the legal obligation to be able to show the process of getting to a policy which would not be possible while using these technologies.

Another finding from the interviews has to do with the acquisition of generative AI contracts. Multiple interviewees note that the contacts are long and hard to read through and don't mention until the last parts that the owners of the models often have the right to train their model using the data sent in. While these suppliers might offer an exception, it became clear that this exception is not always granted.

Besides the fact that the contracts are tough to get through and that not all organisation may bet the exception another aspect of generative AI the constituents struggle with is the investigation into these tools. They notice that their IT-administrators are not always able to accurately determine the risk of enabling these tools. Besides this, it is not always clear that certain updates enable these new features, leading to a sort of feature creep where suddenly it is active.

Besides the IT-administrators, investigations into these models are very hard, the term black box models was mentioned quite a few times. The constituents cannot accurately assess the origin of the data, both the training data and the data that is generated. New risk management procedures are therefore needed for applications which have these features. All interviews agreed with this, some however had not yet been able to set up these new processes.

The results from the interview with the NCSC-NL advisor lead to some new points. Their idea of the cyber resilience of the constituents was relatively accurate, maybe a little lower than the results from the survey suggests. The knowledge of what they thought the constituents rate themselves or their knowledge was a good thing to see. Of the possible publications the interviewees would like to see, they brought up a couple of points that indeed are requested. The alignment of the advisor with the constituents does not appear to be a cause of the research-practice gap within the NCSC-NL.

When asked about this gap however, the advisor noted that it is present within the NCSC-NL. They agreed with some of the theoretical background as for the cause of it, they mentioned the differing types of knowledge and specifically the timeframe making it hard to algin the advisors and the research staff. They did think that having a separate scientific research team did help in making the research-practice gap not as large as it could be if this team was not present. Yet even with the team present and within the same unit, scientific sources are not

being used all that often when writing the publications. This is a confirmation of the point made by Lawler and Benson [21] where they note that the results might not be particularly applicable. The advisors do use more scientific research in the case of generative AI. This is largely because this is a new technology with a lot of hype surrounding it. The findings from researchers are rigorous and provide an insight in what is possible and might be prevalent later.

Besides this, the focus of the publication is often on the here and now both in information the information they provide as well as the steps constituents should take. In an emerging technology, this might not be enough. Some publications have long lead times, meaning that by the time they are finished the subject of the publication might be out of date, involving research papers, like they do with generative AI, might help with being ahead of the curve.

The advisor noted that scientific research in cyber security is practical and therefore could be used as a source for a publication. A lot of the publications however revolve around risk management, and the frameworks the scientific works describe are interesting, however often not used in practice. This sounds like a different research-practice gap, however not relevant to the one in this research.

A final thing to note from the interview with the advisor is that the constituents are not being involved in the publication process. The advisors do maintain a connection with them in other ways so they know what is going on within the constituents but they are not utilised a lot within the publication process. Looking back at the theoretical background, this does provide a great option to involve in the model.

5. Model

Following the results of the survey, the interviews and literature a model has been created. The model consists of steps to take to first identify and second try to mitigate the research-practice gap. The definition of the research-practice gap is "*The disconnect between information researchers publish and what practitioners can use.”*. The research side of this gap is the results of scientific research, research done with much rigor and sometimes difficult language, as mentioned in the theoretical background. The practitioners are the people in the field, the people actively using or dealing with the technology, in this case cyber security professionals. The interviews with the practitioners confirmed that such a gap exists.

In this case the gap is not really on the technical side like how generative AI works, this is known by both sides, it is however on some of the properties of these systems. The main examples of this from the interviews were explainability and trustworthiness of the information.

The goal of this model is to help NCSC-NL in increasing the cyber resilience of its constituents by informing them on time about emerging technologies. The main method to achieve this is shrinking the research-practice gap on emerging technologies. More specifically, the model is focussed on technologies with capabilities that can have an impact on cyber security. In this model, the research part of the gap is used as a guideline to determine what might happen in the future. From the interview with the NCSC-NL advisor it became clear that usually scientific cyber security research is practical and applicable. This means that while not every capability theorised in research will become usable or an attack vector, it is expected that plenty of capabilities will be usable in the future.

An important part of the model is ensuring that information arrives on time. Speed is important in cyber security as it takes a long time to identify and respond to cyber security incidents [46]. Using research as a guide for the future might help in making sure NCSC-NL informs its constituents in a timely manner. Waiting on reports from practitioners to start on a publication will by definition mean that attacks are already happening using the capability. The constituents are then only informed when situations are happening, instead of being able to take preventative measures.

To achieve this timeliness, signposts of change will be used. Signposts of change are a technique used by the intelligence community. It is explained well in [42], in short it is a list

of indicators of a greater topic, that also specify if a certain indicator has been spotted yet. It has been chosen as it is a method which is clear, gives insight on the first look and divides up the capabilities which allows the advisors to more easily make a broad planning. An example of such a signpost can be found in Figure 7. The way to work with this technique is to first choose a topic, in the example this is LLMs. Afterwards, brainstorm ideas on what capabilities this topic will bring. This capability is then split further into indicators which will show the progress being made for the capability to become reality.

The reason for choosing a checklist in combination with signposts is that Chen et al. [32] suggest to balance model accuracy and simplicity. A checklist is a simple method and will help as all steps are easy to follow. Besides this the signposts provide a clear overview of the progress and provide transparency. Transparency was also a best-practice and was found in Sutherland et al. [31]. Transparency was kept central in the designing of this model so constituents trust the results more and allow them to have influence in the process.

Besides timeliness the signposts also help with providing only relevant information to the constituents. If NCSC-NL would write a publication of every scientific paper they read they might overload the constituents with some irrelevant information. Only writing the publications later when certain signposts have been observed helps in only publishing information that is relevant as is recommended by Chen et al. [32].

The reasons for writing a publication by the NCSC-NL instead of for example spreading the knowledge from a research papers are the following, *i)* publications by the NCSC-NL are focussed on Dutch companies and are written in Dutch, *ii)* the knowledge required for practitioners can be different then what is named in a research paper *iii)* the goal of a NCSC-NL publication is not just to inform but also provide some guidance as to what to do next, this usually is lacking in research papers. These reasons also link back to research question 3 which is about what type of information the constituents are missing. The reasons were derived from a combination of both types of interviews. From the interviews with the practitioners it became clear that besides just information, they also wanted to get more knowledge and some wisdom. They wanted frameworks for how to use and implement generative AI safely and knowledge on specific topics like shadow IT. In conclusion it looks like they want actionable advice and frameworks, so information that is higher in the DIKW pyramid [33]. Besides this, it also helps address the point brought up by Lawler and Benson

[21] that the communication style of researchers and practitioners is not the same. NCSC-NL publications are written in the style of the practitioners.

Besides this the NCSC-NL also already does some outreach to the constituents, the way they do this is quite similar to the CEO as described in Lawler and Benson [21]. NCSC-NL goes to the constituents and asks them to sketch an image of what they think the threats are at the moment. This information is then also used to inform the NCSC-NL advisors on what the constituents think is important at the time. Besides this the advisors also provide advice on a case by case basis to constituents. That is also a method of gaining understanding in the goings on and knowledge of constituents.

Steps the NCSC-NL should take to tackle the research-practice gap

Information gathering phase:

1 Monitor scientific papers for new use of the technology with impact on cyber security.

2 Periodically check in with constituents to see opinions and usage of the technology.

3 Align perspectives and look for differences, these differences are your gap.

Preparation phase:

1 Meet with other teams internally to discuss what new or improved capabilities this technology can bring, align the opinions on the capabilities and make a risk assessment per capability

2 Set up signposts of change for every new capability

3 While taking into account risk assessment, determine at what sign what steps in the publication making have to be made.

4 Determine what the publication will be about, its goal, what constituents to involve, the target audience, the timeframe they need to inform the constituents and what steps need to happen at what indicator and observation combination .

Active phase:

1 While taking into account risk assessment, monitor signpost and determine what type of publication will lessen the risk profile of the new capability
2 During check-ins with constituents ask them about the capabilities , their applications and exposure to this capability. Also share NCSC-NL point of view.
3 Actively look for papers based on these capabilities and assess if these change either the signposts or risk assessment.

Publication phase:

1 Have the selected constituents do one last proof read of the publication
2 Adjust the publication to their needs and possibly prepare 2 publications if the needs of both groups are too different.
3 Publish to constituents and possibly the website

5.1 Information gathering phase

The model begins with an information gathering phase. This phase is not a onetime occurrence but recurring. In this phase, the main focus is identifying the gap between the research and practice. This phase came about from the best-practices from the theoretical background. The main point made by the papers in there is that proper communication can help a lot in identifying and shrinking the gap. Knowing what the perspectives of this capability both sides have is the main point in identifying this gap. Finding the gap is done by aligning the perspectives of the scientific research and the practitioners. The differences found are the gap. This gap does not only exist from research to practitioners, but also from practitioners to research. Practitioners might be fearing other parts of the technologies which are not being investigated as much. An example of this came up in the interviews, shadow IT was mentioned a lot while it is not being mentioned much in the same breath as generative AI in a scientific context. This phase is similar to the research setup used for this research. It helps with identifying the gap present and the both NCSC-NL and its constituents were open to participate in it. This section only applies the vision and number of priorities best-practices found in the literature. These are mostly used in step 3 as aligning the perspectives create the vision as well as the number of priorities.

Besides perspectives from both sides one of the main points is for the advisors to ask constituents about the questions they have. If these questions can already be found in the research then maybe it is possible to already answer some of these questions. The questions from the research can mostly be found through looking at the conclusions and discussions, there they propose new research ideas from the results they found.

In short, the gap in this state is the difference in the perspectives and questions between the constituents and research. It is the focus of the publication, it might slightly change during other phases.

5.2 Preparation phase

Once the gap has been identified the next phase beings. This preparation phase is an internal process, it does not have any dependencies on external parties, just NCSC-NL internal ones. Per gap, discuss with other internal teams what they think the capabilities of the technology are and make a risk assessment per capability. During this time, also set up signposts to monitor the progress of the capabilities. An example of a signpost of change is:

Signpost for LLM		
Capability	**Indicator**	**Observation**
Highly personalised LLM assisted spam campaigns	Increased amount of spam	Highly prevalent
	Increased quality of spam	Observed
	Increased quantity of spam domains	Not observed
	Increased personalisation of spam	Not observed

Figure 7: Example of signpost of change on the topic of LLMs

When such a signpost is finished, the advisors themselves should then determine themselves roughly what timeframe they think to need to inform the constituents on this capability. Keeping the risk assessment and this timeframe in mind, they should then plan on what needs to happen at what indicator and capability level, for example if the expectation is that the time between *"Increased amount of spam"* is highly prevalent and highly personalised LLM assisted spam campaigns is expected to be a relatively long time, then maybe the only step that needs to happen is reaching out to the chosen constituents for a check in, if the development of the technology or capability is expected to be fast, then maybe other steps like starting with the publication need to be set. Another goal could be to have the publication

75% done by the time 3/4 indicators have been observed. When choosing the constituents, do keep in mind that later in the process there is a division between constituents who are more ahead of the curve and those who aren't. Both will be asked to proof-read and provide feedback, so when selecting the constituents to involve with this process, keep that in mind too.

A thing to keep in mind when selecting the goal is that if you do mention it in the publication, ensure that it is tested a lot beforehand. A big point of feedback in the survey was that one of the goals set by the reference publication [15] was achieved, or the aspects mentioned. This would seem sloppy and affect the opinion of the NCSC-NL publications.

The reason for using the signposts of change is that it slices up the capabilities into different parts. Each of these parts can be tracked easier. Besides that, it enables the parts of the publication process to be divided. This might help the advisors to ensure that the publication is on time before big incidents using this technologies capability arise.

The following points at least need to be present: *Determine what information the constituents think they need on the capability, determine what the NCSC-NL thinks the constituents need, publication needs to be ready to publish, publication phase starts* and *constituents have proof-read and given feedback on the publication.*

In this phase the scope of the project is defined through setting up the signposts. Scoping is a best-practice found in the literature. Besides this the composition and number of participants and the team are chosen. These are also best-practices. Steps 3 and 4 are similar to steps taken by the NBS to make research applicable.

5.3 Active phase

While the active phase is largely the same as the information gathering phase, the reason for making it a different phase is that it has a different purpose. It is still to identify a gap, but more particularly identify changes in this gap as time moves on. The main addition updating and monitoring of the signposts. This should be slightly easier than before as the technology is now also split up in capabilities, finding and reading papers can now be on those capabilities as well.

This phase is mostly focussed on the best-practices of participant ownership and transparency. Choosing the right type of publication is already used by NCSC-NL when they choose what publication to make. Step 2 could also be considered to be a facilitation of a

breakout session. Using the right advisor for this is important, the best choice is probably the advisor writing the publication as all advisors at least have some experience hosting those kinds of sessions. Participant ownership is stimulated by step 2 as well. By involving the constituents in the data gathering will make them more invested in the product. Transparency is stimulated by sharing the point of view of NCSC-NL as well as having multiple constituents present in in the check-ins.

5.4 Publication phase

The publication phase does not necessarily start after the Active phase, but can occur simultaneously. Actively involve constituents who are ahead of the curve, and people who are behind the curve in two different sessions. In these sessions, first again ask their opinion of the capability and what information they think they need. After that, let them in on the information the NCSC-NL has and ask if the information need changes. Take both into account and adjust the publication. A good thing to note as found from the survey is that the aspects the constituents were questions about looked to correlate well with the eventual end prediction. Using these aspects to test the written products before publication will help in determining the perception after publication.

The publication phase only uses transparency and participant ownership with regards to best-practices. It embraces transparency through having the publication proof read and possibly publishing to NCSC-NL's website as well. Participant ownership is also used by having the constituents proof read the publication. Using the constituents in such a way is also a bit similar to both the CEO from Lawler and Benson [21], as well as the NBS from Bansal et al. [20].

6. Discussion

Having presented the theoretical base, methodology results and the resulting model now the research questions will be answered.

1. How do constituents perceive the reference publication in terms of i) clarity and ii) usefulness?

The constituents gave scores ranging from 3.20 to 4.07 on the questions relating to the reference publication. See Table 2 for the average scores per aspect. These scores were close to the rating of the publication. This suggests that the chosen aspects together provide a good estimation of the rating of the publication. The clarity and usefulness specifically were rated at 3.83 and 3.46 respectively. The constituents thus perceive the reference publication pretty well in terms of clarity and usefulness.

2. How accurately can constituents determine their cyber resilience with regards to generative AI?

When looking at Figure 6 in combination with the constituents own grading of the risk management and cyber security a limitation became clear. Figure 6 is a colour coded table of the answers of the survey on the knowledge of generative AI technologies. There is no conclusion to be drawn from the results of Figure 6. There is no clear correlation between the results of the figure and the grading the constituents gave themselves. Because of this, no real conclusion can be drawn from this data. Researching this can be a good area of future work. To be able to draw a conclusion there probably need to be more respondents and the questions would have to be split up further. The option *"The knowledge is present within the organisation"* could be replaced by a second set of questions of the knowledge of the organisation instead. In addition to this more technologies can be added to present a more nuanced image. Making those alterations to the research setup will increase the likelihood of a conclusion being able to be drawn.

3. What type of information with regards to generative AI are the constituents missing or lacking and how will this information improve the cyber resilience of the constituents?

From the survey it became clear that some technologies are not well known within the constituents. Looking more at the results of the in-depth interviews it became clear that while

those technologies are unknown, that is not where the main questions lie. The main questions were related to risks, implementation, shadow IT and data security. Adding a little context from the advisor interview this is likely because the constituents are mostly worried about the here and now and that if a technology is not widely known now they are not too worried about it. When looking at the information wishes of the constituents and scaling them on the DIKW pyramid it looks like the information need is focussed on the higher level information. The information that will help answer their questions are more frameworks and information rather than intelligence/data. Having access to the information they are looking for will increase the cyber resilience by knowing the risks of the usage of generative AI systems better. This knowledge will help them put in place strategies with regards to the usage generative AI.

In conclusion, the results suggest that the information needed is frameworks from NCSC-NL which will increase the cyber resilience as this information will help the constituents put in place strategies to mitigate these risks.

4. What are the common denominators with regards to the missing information from the NCSC-NL within or across sectors or types of organisations?

After analysis of the data collected by the survey it became clear that no groupings were able to be made. While section 1 of the survey provided a good range of data from the companies, there was no correlation between these data points and other answers of the survey.

When looking at the results from the in-depth interviews all interviewees displayed the need for information and most agreed that data security and risks were important. There was no combination of factors that suggested that all companies needed the same information.

The combination of these data point unfortunately did not give any more insight in common denominators with regards to organisational information and information wish. The wish for frameworks from the NCSC-NL seemed to be universal.

5. What adaptations need to be made to make the best-practices from other disciplines work at NCSC-NL?

Implementing the best-practices from other disciplines appeared quite easy. This was tested by implementing the best-practices into the model design. Implementing these best-practices

was easy as they were quite broad and seemed to fit into the current NCSC-NL processes. Perhaps if the model is validated in future work it would be good also be good to test if the implementations of the best-practices helped in bridging the gap and if they were implemented correctly.

To conclude for this research, it appears that no adaptations need to be made to make best-practices from other disciplines work at NCSC-NL.

How can the challenges of NCSC-NL, regarding their statutory task to increase digital resilience be characterised, when looking at the case of generative AI?

When looking at generative AI it becomes clear from the answers of the interviewees that a research practice gap exists. The constituents have not yet seen incidents that as far as they know made use of generative AI. When looking at the research however there are many different examples of generative AI capabilities and this technology is rapidly evolving. Choosing exactly what information to provide the constituents with is a challenge as information should be timely. Overloading the constituents will also not help so there needs to be some process of choosing the right topics to publish on.

To help NCSC-NL in performing their statutory task a model was created. This model is aimed at identifying and bridging the research-practice gap as well as provide a framework for choosing capabilities to write publications on.

The challenges can be described as having little knowledge of the exact cyber resilience of the constituents, having no clear picture of the information need of the constituents and a difficulty keeping track of the usage of new technologies in practice.

6.2 Limitations

As briefly discussed in the results, the replies by sector and replies by role paint important context for this thesis. The distribution was set through the responses on the survey. Since this distribution is not every company, or every role within a company the results are skewed towards the knowledge and needs of the respondents. An example of this is that most of the interviews with people who are high up in the company. The information they needed also made a lot of sense for their position, however the information they need is different to someone who works more in the trenches of cyber security. Unfortunately this bias in the data is present whenever not everyone responds to the survey. The spread however does provide a wide sample, a lot of roles and sectors were surveyed. For the purpose of seeing if there is,

and possibly measuring a gap the distribution works. It has showed that the constituents that replied are relatively alike in their need and want for information and that there is no couple of factors that can describe them all.

In addition to possible bias, it is also worth noting that 14 out of the 30 respondents were CISO/ISO's. While this might again bring in a little more bias, these are the roles that a decent amount of the publications are focussed towards. This is because those publications revolve around risk management which is usually handled by the CISO/ISO. Those roles are therefore important stakeholders for the NCSC-NL, it is good to see they responded as much as they have.

Unfortunately the sample size and results combination lead to us not being able to answer research question 4. There did not appear to be any common denominators when it comes to the results. In addition to this, the setup might not have been perfect to answer these questions. In the open questions in the survey this was asked, however the main point of information was the interviews. The interviews were only held with 5 people so unfortunately a real common denominator is not possible to point out there. While most people did agree on the fact that they need more information about generative AI and shadow IT, this was said by all 5 interviewees, who all have different backgrounds. In future research this is something to keep an eye out for.

In the selection process for the technologies there was a deliberate choice to include two types of technologies. Some technologies were expected to score quite high, like Gemini, prompt injection and deepfakes, whereas others were expected to have a low knowledge grade like the open source models Mistral 22B and LLaMa 2. This was largely to expectation as these were deemed to be known techniques when it comes to generative AI. The amount of information being present in the organisations positively surprised us, however the amount of unknown answers equally surprised us. For a next survey, a couple more techniques could be used, as well as possibly asking both the individual on a 5 point scale, as well as the organisation. Some CISO's might not know everything, however it is possible that within the organisation there are projects being done and experts being hired. The reason for asking about these techniques was to see if it relates at all to how cyber resilient the constituents think they are. Unfortunately the result of this is inconclusive. There is no link between the grades of their risk management or cyber security and the responses on the technologies. This means that for research question 2 the answer is also inconclusive. That does however

provide a good opportunity for new research, investigating how accurately constituents can gauge their cyber resilience.

The reference publication [15] was not the first choice of publication. At first a NCSC-NL publication was supposed to be attached. This was chosen to increase the response rate as well as test a NCSC-NL publication directly. Because of time constraints the research pivoted to using the AISC reference publication as the NCSC-NL publication would not be done on time. To answer the research questions some changes to the survey were made to make the results still applicable and to still be able to answer the research questions. For future work, if the time is available, sending the survey out with a NCSC-NL publication might lead to slightly different or more applicable results.

The main limitation of the model is that it is not validated. While it was in the timeline to have a third party give their opinion on the model as validation, because of time considerations this was not done. This means that it is inconclusive if the model works and will work for the purpose it is designed. This is perhaps an area of future work. The third party would have been an expert from an organisation similar to NCSC-NL. This expert is an advisor in the other organisation and has a PhD in AI. This person is qualified as they can determine whether or not the usage of the model will aid in addressing the research-practice gap through their experience and knowledge.

7. Conclusion

In the theoretical background to the right definition and best-practices for the research-practice gap were presented. After this the research setup was discussed in the methodology. This was followed up by the presentation of the results in the results section. A model based on the findings of the results and literature was presented in the model chapter. The research questions brought up in the introduction were then reflected upon in the discussion as well as a presentation of the limitation of this research.

In this research it was found that NCSC-NL is dealing with a research-practice gap when it comes to information about generative AI. The findings of this research might not be not fully representative as only 30 constituents filled in the survey. There was however width in the organisations, sectors and roles within the organisations. This leads us to believe that there is enough base to characterise the gap. This gap mostly revolves around capabilities displayed in research and what is visible to practitioners. Besides this practitioners also had many other questions that did not come back in the research.

To help bridge this gap a model was presented. This model is based on the findings of the survey and interviews as well as the best-practices found in the literature. The model is a checklist in combination with signposts of change. We propose that NCSC-NL use this model to help bridge the gap to inform its constituents on the topic of generative AI.

Bibliography

[1] S. Houde, V. Liao, J. Martino, M. Muller, D. Piorkowski, J. Richards, J. Weisz and Y. Zhang, “Business (mis)Use Cases of Generative AI,” *http://arxiv.org/abs/2003.07679,* 2020.

[2] H. Naveed, A. U. Khan, S. Qiu, M. Saqib, N. Akhtar, N. Barnes and A. Mian, “A Comprehensive Overview of Large Language Models,” *http://arxiv.org/abs/2307.06435,* 2024.

[3] CNN, “Finance worker pays out $25 million after video call with deepfake ‘chief financial officer’,” CNN, 04 February 2024. [Online]. Available: https://edition.cnn.com/2024/02/04/asia/deepfake-cfo-scam-hong-kong-intl-hnk/index.html. [Accessed 22 08 2024].

[4] T. Pung, V.-A. Pădurean, J. Cambronero, S. Gulwani, T. Kohn, R. Majumdar, A. Singla and G. Soares, “Generative AI for Programming Education: Benchmarking ChatGPT, GPT-4, and Human Tutors,” in *ICER 2023: ACM Conference on International Computing Education Research,* Chicago, 2023.

[5] Z. Yang, B. Xu, J. M. Zhang, H. J. Kang, J. Shi, J. He and D. Lo, “Stealthy Backdoor Attack for Code Models,” *http://arxiv.org/abs/2301.02496,* 2023.

[6] R. Fang, R. Bindu, A. Gupta, Z. Qiusi and D. Kang, “LLM Agents can Autonomously Hack Websites,” *https://arxiv.org/abs/2402.06664*, 2024.

[7] OpenAI, “ChatGPT release notes,” OpenAI, 15 12 2022. [Online]. Available: https://help.openai.com/en/articles/6825453-chatgpt-release-notes. [Accessed 21 04 2024].

[8] Open AI, “New models and developer products announced at DevDay,” 06 11 2023. [Online]. Available: https://openai.com/blog/new-models-and-developer-products-announced-at-devday. [Accessed 22 08 2024].

[9] Rijksoverheid, “Wet beveiliging netwerk- en Informatiesystemen (WBNI),” 01 09 2018.

[Online]. Available: https://www.rijksoverheid.nl/documenten/rapporten/2018/09/01/wet-beveiliging-netwerk--en-informatiesystemen-wbni-voor-digitale-dienstverleners. [Accessed 03 06 2024].

[10] NCSC, "Statutory Task," [Online]. Available: https://english.ncsc.nl/about-the-ncsc/statutory-task. [Accessed 03 06 2024].

[11] Europa.eu, "NIS Directive," 2024. [Online]. Available: https://www.enisa.europa.eu/topics/cybersecurity-policy/nis-directive-new. [Accessed 03 06 2024].

[12] NIST, "Cyber resiliency," NIST, [Online]. Available: https://csrc.nist.gov/glossary/term/cyber_resiliency. [Accessed 22 08 2024].

[13] J. Zhang, H. Wen, L. Deng, M. Xin, Z. Li, L. Li, H. Zhu and L. Sun, "HackMentor: Fine-Tuning Large Language Models for Cybersecurity," in *2023 IEEE 22nd International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*, Exeter, 2023.

[14] O. Tkachenko, H.-J. Hahn and S. L. Peterson, "Research–Practice Gap in Applied Fields: An Integrative Literature Review," *Human Resource Development Review,* vol. 16, no. 2, pp. 235-262, 2017.

[15] AISC, CISA, FBI, ACSC, CCCS, NCSC-NZ, NCSC-UK, "Deploying AI Systems Securely," 15 03 2024. [Online]. Available: https://www.nsa.gov/Press-Room/Press-Releases-Statements/Press-Release-View/Article/3741371/nsa-publishes-guidance-for-strengthening-ai-system-security/. [Accessed 22 08 2024].

[16] R. Hall and P. Andriani, "Managing Knowledge for Innovation," *Long Range Planning,* vol. 35, no. 1, pp. 29-40, 2002.

[17] K. A. Robinson, I. J. Saldana and N. A. Mckoy, "Development of a framework to identify research gaps from systematic reviews," *Journal of Clinical Epidemiology,* vol. 64, no. 12, pp. 1325-1330, 2011.

[18] J. M. Bartunek and S. L. Rynes, "Academics and Practitioners Are Alike and Unlike:

The Paradoxes of Academic–Practitioner Relationships," *Journal of Management,* vol. 40, no. 5, pp. 1181-1201, 2014.

[19] K. Starkey and P. Madan, "Bridging the Relevance Gap: Aligning Stakeholders in the Future of Management Research," *British Journal of Management,* vol. 12, no. s1, 2001.

[20] P. Bansal, S. Bertels, T. Ewart, P. MacConnachie and J. O'Brien, "Bridging the Research-Practice Gap," *Academy of Management Perspectives,* vol. 26, no. 1, pp. 73-92, 2012.

[21] E. E. Lawler and G. S. Benson, "The practitioner-academic gap: A view from the middle," *Human Resource Management Review,* vol. 32, no. 1, p. 100748, 2022.

[22] A. Kieser and L. Leiner, "Why the Rigour–Relevance Gap in Management Research Is Unbridgeable," *Journal of Management Studies,* vol. 46, no. 3, pp. 516-533, 2009.

[23] A. H. Van De Ven and P. E. Johnson, "Knowledge for theory and practice.," *The Academy of Management Review,* vol. 31, no. 4, pp. 802-821, 2006.

[24] R. Khurana, Rakesh Khurana: From Higher Aims to Hired Hands: The Social Transformation of American Business Schools and the Unfulfilled Promise of Management as a Profession, Princeton: Princeton University Press, 2007.

[25] M. Kelemen and P. Bansal, "The Conventions of Management Research and their Relevance to Management Practice," *British Journal of Management,* vol. 13, no. 2, pp. 97-108, 2002.

[26] P. Shrivasta and I. I. Mitroff, "Enhancing Organizational Research Utilization: The Role of Decision Makers' Assumptions," *The Academy of Management Review,* vol. 9, no. 1, p. 18, 1984.

[27] KPMG, "Evidence-based HR: The bridge between your people and delivering business strategy," KPMG International, 2015.

[28] D. McIntyre, "Bridging the gap between research and practice," *Cambridge Journal of Education,* vol. 35, no. 3, pp. 357-382, 2005.

[29] J. Roloff, "Learning from Multi-Stakeholder Networks: Issue-Focussed Stakeholder Management," *Journal of Business Ethics,* vol. 82, no. 1, pp. 233-250, 2008.

[30] S. L. Rynes, T. L. Giluk and K. G. Brown, “The Very Separate Worlds of Academic and Practitioner Periodicals in Human Resource Management: Implications for Evidence-Based Management,” *Academy of Management Journal,* vol. 50, no. 5, pp. 987-1008, 2007.

[31] W. J. Sutherland, E. Fleischman, M. B. Mascia, J. Pretty and M. A. Rudd, “Methods for collaboratively identifying research priorities and emerging issues in science and policy,” *Methods in Ecology and Evolution,* vol. 2, no. 3, pp. 238-247, 2011.

[32] X. Chen, T. Deng, Z.-j. M. Shen and Y. Yu, “Mind the gap between research and practice in operations management,” *IISE Transactions,* vol. 55, no. 1, pp. 32-43, 2023.

[33] J. Rowley, “The wisdom hierarchy: representations of the DIKW hierarchy,” *Journal of Information Science,* vol. 33, no. 2, pp. 163-180, 2007.

[34] F. Hare, “Private Sector Contributions to National Cyber Security: A Preliminary Analysis,” *Journal of Homeland Security and Emergency Management,* vol. 6, no. 1, 2009.

[35] M. Carr, “Public-private partnerships in national cyber-security strategies,” *International Affairs,* vol. 92, no. 1, pp. 43-62, 2016.

[36] K. Stoddart, “UK cyber security and critical national infrastructure protection,” *International Affairs,* vol. 92, no. 5, pp. 1079-1105, 2016.

[37] D. Story and A. Tait, “Survey Research,” *Anesthesiology,* vol. 130, no. 2, pp. 192-202, 2019.

[38] A. Bryman, “Integrating quantitative and qualitative research: how is it done?,” *Qualitative Research,* vol. 6, no. 1, pp. 97-113, 2006.

[39] G. D. Rubenfeld, “Surveys: an introduction,” *Respiratory Care,* vol. 49, no. 10, pp. 1181-1185, 2005.

[40] C. Wilson, “Interview Techniques for UX Practitioners,” in *Interview Techniques for UX Practitioners,* Elsevier, 2014.

[41] M. Kiger and L. Varpio, “Thematic analysis of qualitative data: AMEE Guide No. 131,”

Medical Teacher, vol. 42, no. 8, pp. 846-854, 2020.

[42] CIA, A Tradecraft Primer, US Government, 2009.

[43] A. J. Neumann, S. N. and R. D. Webb, “Post processing audit tools and techniques,” NIST, 1977.

[44] K. Greshake, S. Abdelnabi, S. Mishra, C. Endres, T. Holz and M. Fritz, “Not What You've Signed Up For: Compromising Real-World LLM-Integrated Applications with Indirect Prompt Injection,” in *CCS '23: ACM SIGSAC Conference on Computer and Communications Security*, Copenhagen, 2023.

[45] Microsoft, “Staying ahead of threat actors in the age of AI,” Microsoft, 14 02 2024. [Online]. Available: https://www.microsoft.com/en-us/security/blog/2024/02/14/staying-ahead-of-threat-actors-in-the-age-of-ai/. [Accessed 08 08 2024].

[46] Forbes, “The Importance Of Time And Speed In Cybersecurity,” Forbes, 22 01 2021. [Online]. Available: https://www.forbes.com/sites/forbestechcouncil/2021/01/22/the-importance-of-time-and-speed-in-cybersecurity/. [Accessed 22 08 2024].

[47] Cisco, “Cisco AI Readiness Index,” Cisco, 2023. [Online]. Available: https://www.cisco.com/c/m/en_us/solutions/ai/readiness-index.html . [Accessed 03 04 2024].

[48] J. Jöhnk, M. Weißert and K. Wyrtki, “Ready or Not, AI Comes— An Interview Study of Organizational AI Readiness Factors,” 2021.

[49] D. M. Rousseau, J. Manning and D. Dener, “Evidence in Management and Organizational Science: Assembling the Field’s Full Weight of Scientific Knowledge Through Syntheses,” 2008.

Appendix

A. Survey questions

The original survey questions are in Dutch, this is a translated version.

Dear NCSC-NL-partner,

Thank you for taking a look at this survey. The expected time to answer all the questions is 5-7 minutes.

All questions are about your own thoughts or knowledge and are not too hard to answer. Some questions are about the added publication about generative AI, please read this before answering the questions about it.

All data is stored on NCSC-NL systems and won't be shared with any third parties. The data will be deleted after the research is concluded and will be used anonymized.

Section 1/5

What is the name of your organisation?

What is the size of your organisation? (MC)

1-50 51-250 251-1000 1001-5000 5000+

What is your role at the organisation?(MC)

SOC employee – Risk manager – CISO – IT-administrator – ISO – Other:

What is the revenues/budget of your organisation? (MC)

< 1.000.000 < 10.000.000 < 50.000.000 < 100.000.000 < 1.000.000.000 > 1.000.000.000

Roughly what is the IT-budget of your organisation? (MC)

< 1.000.000 < 10.000.000 < 50.000.000 < 100.000.000 < 1.000.000.000 > 1.000.000.000

How big is the IT-security team in FTE? (MC)

0 1-5 5-25 25-50 50+

Does your organisation have a CIO? (MC)

Yes No

Does your organisation have a CISO? (MC)

Yes No

Does your organisation have a SOC? (MC)

Yes No

In what sector is your organisation active? (MC)

Energy – Telecommunications – Transport – Drinking water – Water – Chemistry – Nuclear – Finance – Central government – Decentral government – Other:

The following sections are about the publication. Please read the publication before answering the questions.

Rate on a scale from 1-5 how much you agree with the following statements. Where 1 Is completely disagree and 5 is completely agree.

Section 2/5

The technical level of the information was right.

The publication is action oriented.

The publication is useful.

The publication is clear.

The publication is informative.

The information in the publication is complete.

How many stars would you rate this publication (1 = bad, 5 = good).

[Open field for motivation]

In this section you will be asked more questions about your opinion of the publication as well as how you would usually use NCSC-NL publications.

Section 3/5

The publication notes that it wants to improve the CIA principles, on a scale from 1-5 rate how well the publication does this.

The publication notes that it wants to ensure that known vulnerabilities in AI systems are mitigated the right way, on a scale from 1-5 rate how well the publication does this.

The publication notes that it wants to provide methodologies and controls to protect, detect, and respond to malicious activity against AI systems and related data and services, on a scale from 1-5 rate how well the publication does this.

Why did or didn't the publication achieve its goal?

What is the target audience for this publication?

Low IT maturity organisations – Average IT maturity organisations – High IT maturity organisations.

What is the target sector for this publication?

Energy – Telecommunications – Transport – Drinking water – Water – Chemistry – Nuclear – Finance – Central government – Decentral government – Other:

What is the target audience for this publication within those organisations?

SOC employee – Risk manager – CISO – IT-administrator – ISO – Other:

In the next section, some technologies will be named and we will ask how familiar you are with these technologies.

Section 4/5

The options in this section for all techniques are:

Unknown - I looked it up once - The knowledge is present in the organisation - I could apply it - I am an expert at it

The techniques are:

Prompt injection

Automated hacking

Deepfakes

LLM finetuning

Model Stealing

Vector databases

Mistral 22b

LLaMa2

Gemini

Notes (optional)

In this next section you will be asked about your opinion on the maturity and generative AI processes within your organisation.

This information will only be used for this research, this data will not be shared. is stored on NCSC-NL systems and will be deleted when the research is concluded.

If you do not feel comfortable answering the first 2 questions, choose N/A.

Section 5/5

On a scale from 1-5, rate the maturity of the risk management at your organisation.

On a scale from 1-5, rate the maturity of the cyber security at your organisation.

To your knowledge, how many generative AI applications are in use internally.

0 1-5 5-25 25-50 50+

To your knowledge, what are the generative AI applications?

To your knowledge, what departments make use of generative AI applications?

Are there any limitations on the usage of generative AI applications at your organisation?

Thank you very much for participating in this survey, it will help the research a great deal.

If you would like to receive a management summary of the results and the eventual research paper or if you are open for an in-depth interview following this survey, please leave your email below.

What is your email address?

Are you open for an in-depth interview following this survey?

Yes No

B. Interview guide

The original interview guideline is in Dutch, this is a translated version.

Interview instructions

The introduction will be shared with the interviewee.

The underlined questions are most important. These have to be asked and will be prioritised in the case of a lack of time.

The indented questions are guidelines for further questions.

Introduction

{Introduction of interviewers and interviewee}

My assignment is to advise NCSC-NL on how they should go about generating advice on the case of generative AI. The assumption is that there is a research-practice gap between science and the developers of generative AI and the constituents of NCSC-NL.

If you are good with it, I will record the interview, this is just to make a transcription. The transcription will be deleted at the end of the research period, the recording will be deleted when the transcription is finished. I will not quote from the interview.

The interview will take roughly an hour. There are three topics we would like to cover, generative AI within your organisation, the role of NCSC-NL and your personal opinion about generative AI.

Generative AI within your organisation (15 min)

What is being done with regards to generative AI within your organisation?

> Has there been a noticeable difference in productivity since the adoption of generative AI?

What vision and strategy is held on the topic of generative AI?

> What role does cyber security play in this vision and strategy?

What is the risk-management process like of the adoption of new information systems?

Was this process different for products with generative AI?

What was the biggest obstacle in researching the security of generative AI?

When looking at the security considerations of using generative AI, what aspects are looked at the most?

Training, deployment, possible data leaks?

What step in the development/deployment of generative AI has the biggest security impact?

What models have been looked at or are in use?

If no open source models, why not?

Are you resilient when it comes to generative AI?

Have there been incidents where generative AI have played a role?

Role NCSC-NL (20 mins)

What information do you expect from NCSC-NL when it comes to generative AI?

What information would you like in an ideal world?

Why do you expect this information? Are you missing it now? Are you missing a reputable source?

A lot of organisations admitted to using the TLS standards written by NCSC-NL, what could be the TLS standard of generative AI? (If you could choose 1 publication, what would it be?)

What type of publications from NCSC-NL do you use?

Why these and not others? What do they add?

On what technical level do you expect NCSC-NL publications to be?

What information on the topic of generative AI is missing that NCSC-NL could provide?

What sources do you currently use to follow the developments of generative AI?

Opinion of generative AI (10 min)

How will generative AI change the way of working in 5 years, in your role and in your organisation?

How do you think employees can use generative AI in their job in IT-security?

> Do you use it yourself? Would you like to use it but are you not allowed?

How is generative AI currently making the biggest impact, both in general but also within security?

Personally, what do you think generative AI looks like in the future in general?

> Specifically, what do you think it will look like for cyber security (attacks or defence with generative AI?)

Conclusion

Thank you for your time, it is very much appreciated.

I will share the results when the research is finished, both in a management summary as well as the complete research rapport.

C. Advisor interview guide

What is the most common answer for how well people think they know the following technologies?

Options are:

"Unknown, I looked it up once, The knowledge is present in the organisation, I could apply it and I am an expert at it".

prompt injection, automated hacking, deepfakes, LLM finetuning, model stealing, vector databases, Mistral 22b, LLaMa 2 and Gemini

On average, what grade (1-5) do constituents that participated in the survey their risk management?

On average, what grade (1-5) do constituents that participated in the survey cyber security?

What are the biggest obstacles in investigating generative AI products?

What is the biggest headache of a CISO in investigating generative AI products?

Do constituents think they are cyber resilient?

What information when it comes to generative AI do the constituents want?

What publication would they really like to see?

What is expected of NCSC-NL by the constituents?

How advanced is the usage of generative AI within the constituents?

What percentage of constituents use generative AI?

Do you use scientific papers as a source?

Is scientific research applicable/practical?

Are things like spectre and meltdown applicable?

What is the process of getting to a publication?

Are constituents used in the publication process?

What is the process like after a publication is chosen?

What sources are used?

Are the ideas verified?

Do you have any idea what the research-practice gap is?

What do you think of the research-practice gap?

If there is a research-practice gap, on what side is NCSC-NL?

If NCSC-NL is in towards middle, would it be better?

D. Publication