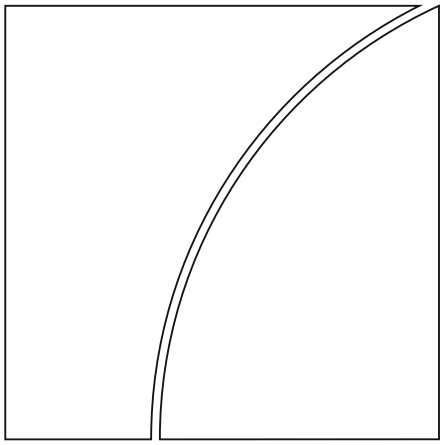


BIS Papers

No 145

Generative artificial intelligence and cyber security in central banking

by Iñaki Aldasoro, Sebastian Doerr, Leonardo Gambacorta, Sukhvir Notra, Tommaso Oliviero and David Whyte

Monetary and Economic Department

May 2024

JEL classification: E58, G20, G28, K24.

Keywords: artificial intelligence, cyber security, central banks, human capital.

The views expressed are those of the authors and not necessarily the views of the BIS.

This publication is available on the BIS website (www.bis.org).

ISSN 1682-7651 (online)
ISBN 978-92-9259-762-7 (online)

Generative artificial intelligence and cyber security in central banking

Iñaki Aldasoro, Sebastian Doerr, Leonardo Gambacorta, Sukhvir Notra, Tommaso Oliviero and David Whyte*

Abstract

Generative artificial intelligence (gen AI) introduces novel opportunities to strengthen central banks' cyber security but also presents new risks. We use data from a unique survey among cyber security experts at major central banks to shed light on these issues. Responses reveal that most central banks have already adopted or plan to adopt gen AI tools in the context of cyber security, as perceived benefits outweigh risks. Experts foresee that AI tools will improve cyber threat detection and reduce response time to cyber attacks. Yet gen AI also increases the risks of social engineering attacks and unauthorised data disclosure. To mitigate these risks and harness the benefits of gen AI, central banks anticipate a need for substantial investments in human capital, especially in staff with expertise in both cyber security and AI programming. Finally, while respondents expect gen AI to automate various tasks, they also expect it to support human experts in other roles, such as oversight of AI models.

JEL classification: E58, G20, G28, K24.

Keywords: artificial intelligence, cyber security, central banks, human capital.

* Iñaki Aldasoro, Sebastian Doerr, Leonardo Gambacorta, Sukhvir Notra and David Whyte: Bank for International Settlements (BIS). Tommaso Oliviero: University of Naples Federico II and Centre for Studies in Economics and Finance (CSEF). We would like to thank Douglas Araujo, Giulio Cornelli, Francesco Paolo Landolfo, Fernando Perez-Cruz and Hyun Song Shin for valuable comments and suggestions. The views expressed in this paper are those of the authors and not necessarily those of the BIS.

1. Introduction

Cyber attacks have become more frequent and sophisticated, and the financial sector consistently ranks as one of the most attacked industries (Aldasoro et al (2020, 2022)). Central banks represent a natural target for cyber attacks, as they are responsible for the management and oversight of critical infrastructures in the financial sector (eg payment systems) and safeguard confidential information about future policy decisions (Doerr et al (2022)). Reinforcing these concerns, in March 2024 a report by the US Department of the Treasury highlighted generative artificial intelligence (gen AI) as an emerging critical aspect for the cyber security of the financial sector.[1]

The emergence of gen AI models, which gained significant momentum with the launch of ChatGPT in late 2022,[2] introduces both opportunities and challenges for the management of cyber risk in the financial sector, including central banks. On the one hand, as gen AI tools become more sophisticated and their use more widespread, the frequency and speed of cyber attacks are set to increase. Such attacks are also likely to become more complex, due to more refined algorithms. Specific threats include AI-generated social engineering, zero-day attacks and malware attacks for data leakage. The adoption of gen AI for internal organisation purposes and potentially also for cyber defence also creates the risk of attacks against AI systems directly. On the other hand, gen AI can strengthen cyber security by enabling the processing of increasingly larger data sets with more sophisticated analytics. It could also help users employ more proactive cyber security and fraud prevention strategies.

In light of these developments, understanding the impact of gen AI on central banks' cyber risk management is of paramount importance. This issue is naturally intertwined with the skills of central bank staff. Indeed, establishing rules of conduct and ensuring a thorough understanding of the risks and benefits associated with the use of gen AI tools for all employees are essential to maintaining high standards of cyber security. Additionally, the question of whether the development of AI tools will complement or replace human expertise in central bank IT units remains the subject of an ongoing debate.

To investigate the link between gen AI and cyber risk we draw on the results of an ad hoc survey conducted among the members of the Global Cyber Resilience Group (GCRG) in January 2024. The GCRG is one of the initiatives of the Cyber Resilience Coordination Centre (CRCC).[3] Established in 2020, the GCRG serves as a forum where chief information security officers (CISOs) from central banks convene to discuss both tactical and strategic issues related to cyber security. The survey gathered responses from 32 participants, delving into the opportunities and

[1] See US Department of the Treasury (2024).

[2] ChatGPT reached 1 million users in less than a week and over 100 million in less than a year.

[3] The CRCC is a business unit at the BIS that leads cyber resilience and collaboration initiatives in the central bank community and comprises experts from 58 BIS member central banks.

challenges that central banks see in the adoption of gen AI tools.[4] Four broad considerations guided the design of the survey:

1) What is the current status of gen AI adoption by central banks?
2) How do central banks evaluate the benefits and challenges for cyber security associated with the use of gen AI?
3) How prepared are central banks for the "AI revolution" and its expected impact on cyber security and data protection?
4) What are the key strategic, ethical and regulatory concerns regarding gen AI adoption in cyber security?

Survey responses reveal four main insights.

First, a large majority of central banks report that they are already using gen AI tools or are planning to do so in the coming years. Respondents indicate that gen AI offers more benefits than risks, especially with regard to specific aspects of cyber security such as cyber threat detection. Yet, the adoption process comes with significant challenges, most notably in terms of adequate investment in human capital. Indeed, more than half of the surveyed experts report that their strategies regarding the evaluation and adoption of AI strategy are currently under development.

Second, the prevailing view is that gen AI can outperform traditional methods in enhancing cyber security management, but that it also introduces new risks. The benefits are largely perceived in specific areas of cyber security, such as automation of routine tasks. AI is expected to reduce the costs associated with time-consuming activities traditionally performed by humans. Additional benefits from gen AI include enhanced threat detection, faster response times to cyber attacks, and learning of new trends, anomalies or correlations that might not be obvious to human analysts. In terms of risks, gen AI can introduce new vulnerabilities into central banks' cyber security defences. Risks related to social engineering and zero-day attacks as well as unauthorised data disclosure are of highest concern.

Third, our results highlight key aspects regarding investments in IT and human capital. Two critical dimensions regarding human capital arise from the survey. The first pertains to all employees at central banks and involves the adoption of gen AI tools, which may be hindered by insufficient technological skills. Indeed, most central banks have enabled or plan to enable their staff to access cloud-based gen AI applications, albeit with certain restrictions on use. This approach aims to mitigate the risks associated with the adoption of gen AI, particularly concerning staff's current unpreparedness for integrating and operationalising AI systems. The second dimension relates to the accumulated human capital in IT divisions at central banks. There is a consensus that gen AI could replace staff in cyber security units for routine tasks. This shift could free up resources to be reallocated towards more strategic cyber security initiatives, potentially increasing productivity. Moreover, responses indicate that even though gen AI is seen as a technology that can handle operational tasks more effectively, it will still require human supervision to ensure ethical and accurate outcomes and continuously train the AI systems. A general concern is thus

[4] The survey includes responses by central banks across both advanced and emerging market economies; the nationality of respondents is anonymised for confidentiality reasons.

the limited availability of personnel with sufficient knowledge of both AI methodologies and cyber security.

Fourth, the consensus among respondents is that gen AI systems will facilitate a shift from a reactive to a proactive approach to predicting and neutralising threats. A critical consideration is the extent of autonomy to be granted to AI tools in cyber security and the nature of their interaction with humans. For an appropriate strategy, data scientists, AI security analysts and AI supervisors are identified as key professional roles for the seamless integration of gen AI with existing security tools.

The results from this study contribute to the ongoing discussion on how to best use AI to limit cyber risk. Kashyap and Wetherilt (2019) propose a set of principles to consider in regulating cyber risk within the financial sector. Moreover, the Basel Committee on Banking Supervision (BCBS) has issued guidelines for banks on best practices regarding cyber risk management (BCBS (2018, 2021)). Our results suggest that additional measures might be necessary to account for the potential benefits and challenges arising from the spread of gen AI.[5] Furthermore, given the significant uncertainty and variability in cost estimates for cyber security incidents – which could increase with the future adoption of AI tools – establishing common guidelines and practices for all central banks is desirable. In addition, existing IT staff may not be fully prepared to handle the fast and disruptive innovations associated with gen AI. The foreseen "skill gap" may be difficult to close for most central banks, given the limited labour supply and high costs associated with new hires.[6] Tackling this issue is vital for central banks going forward.

A growing body of work studies the role and impact of cyber threats within the private sector, including financial institutions and crypto (Boissay et al (2022)), yet few studies have examined implications for central banks. An exception is the work by Doerr et al (2022), who analyse the issue of cyber risk in central banking by leveraging an ad hoc survey conducted in 2021. Our contribution is to broaden the scope of that analysis by enlarging the sample of central banks participating in the survey and incorporating new evidence on cyber security management, specifically regarding the relationship between the advent of new AI tools and the preparedness of central bank staff for this disruptive technological advancement.

Our paper also contributes to the more general debate on the expected impact of introducing gen AI tools on the organisational structure of both public and private companies. The literature highlights an expected rise in labour productivity, especially in tasks that require cognitive abilities (Brynjolfsson et al (2023); Noy and Zhang (2023); Peng et al (2024)), although the effects could be quite different across sectors (Felten et al (2021)). This paper delves into the effects of gen AI on productivity by considering the perspective of cyber security experts in central banks. Our findings underscore that gen AI tools are expected to enhance the efficiency of existing cyber security practices, thereby boosting the productivity of professionals, particularly in

[5] This could include, for example, developing codes of conduct for employees to ensure information protection, as well as promoting the transparency and accountability of AI models, and adherence to international laws.

[6] For example, a 2020 report from the Enterprise Strategy Group (ESG) and Information Systems Security Association (ISSA) revealed that 70% of cyber security professionals indicated that their organisations suffered from a cyber security skills shortage, and more than 60% mentioned that security positions remained vacant for at least three months (see ESG and ISSA (2020)).

routine and operational tasks. At the same time, cyber security experts do not expect their roles to be replaced by gen AI applications; rather, growing importance is placed on human involvement in tasks related to oversight of and training for AI-driven activities, ensuring ethical and accurate outcomes.

The rest of the paper is organised as follows: Section 2 provides context on recent developments in gen AI and how it can affect cyber security. Section 3 presents the survey results concerning the adoption of AI by central banks. Section 4 elaborates on perceived opportunities, risks and challenges associated with the adoption of gen AI tools for cyber security management. Section 5 examines the responses related to questions on investment in IT and human capital. Section 6 discusses the future landscape and provides some regulatory insights. Section 7 concludes.

2. Gen AI and cyber risk

Gen AI can be seen as the latest advancement of machine learning. Broadly speaking, machine learning comprises a set of techniques designed to extract information from data, with a view to making predictions. It can be seen as an outgrowth of traditional statistical and econometric techniques, although it does not rely on a prespecified model or statistical assumptions such as linearity or normality. The process of fitting a machine learning model to data is called training. The criterion for successful training is the ability to predict outcomes on previously unseen ("out of sample") data, irrespective of how the models predict them.

Neural networks are perhaps the most important technique in machine learning, with widespread uses even for the latest generation of models. Their main building blocks are artificial neurons, which take multiple input values and transform them in a non-linear way to output a single number – like logistic regressions. The artificial neurons are organised to form a sequence of layers that can be stacked: the neurons of the first layer take the input data and output an activation value. Subsequent layers then take the output of the previous layer as input, transform it and output another value, and so forth. This way, similar to neurons in the human brain, an artificial neuron's output value is akin to an electrical impulse transmitted to other neurons. A network's depth refers to the number of layers. The weights and biases determining the strength of connections across neurons and layers are collectively called parameters. These parameters are improved iteratively during training. Deeper networks with more parameters require more training data but predict more accurately. Neural networks are behind face recognition or voice assistants like Siri or Alexa and underlie the most significant recent innovations in AI.

Transformers, unveiled in 2017, drastically improved the performance of neural networks in natural language processing (NLP) and enabled the rise of large language models (LLMs). Rather than just relating a word to those near it, transformers attempt to capture the relationship between the different components of a text sequence, even if they are far apart in the paragraph or document. This allows the model to better understand context and hence different meanings a word can have. For example, the meaning of the word "bank" differs when it appears in the sentence "I'll swim across the river to get to the other bank" versus "I crossed the street to go to the bank". Transformers unlocked use cases of natural language processing that

require dealing with long streams of text and gave rise to the most recent advances in LLMs, such as ChatGPT. The availability of huge amounts of digitised text from the internet and rapid advances in computing power have allowed transformer-based LLMs to achieve human-like abilities in processing language.

LLMs underlie the rapid rise of gen AI, which generates content based on suitable prompts and can perform tasks beyond language recognition. In particular, it can generate new content, from text and images to music and code, based on the data they have been trained on.[7] For instance, LLMs like ChatGPT (Generative Pretrained Transformer) are designed to predict the next word or token. By ingesting the Common Crawl in their training set, they present zero-shot learning capabilities and can respond to human prompts to generate text in a way that mimics human language. They thus enable a wide range of linguistic tasks, essay writing, software coding and even engaging in nuanced conversations.

Gen AI is expected to be the next general purpose technology, with the potential to transform various industries and have a significant impact on the economy (McKinsey (2023), Aldasoro et al (2024a)). Central banks are no exception. Indeed, in the recent past, central banks have successfully implemented traditional AI tools for various applications, including machine learning techniques for data analysis, payment systems oversight, supervision and cyber security (Araujo et al (2022, 2024); Doerr et al (2021)).

The adoption of gen AI by central banks provides novel opportunities and challenges related to cyber security management. Gen AI can be applied on both the offensive and the defensive sides of cyber risk (Neupane et al (2023)). In what follows we will report on major risk areas introduced by the offensive use of gen AI models, as well as the implications for cyber defence and the setup of strategies aiming to counter the increase in gen AI-induced cyber threats.

The development of gen AI tools could enhance the capabilities of sophisticated cyber threat actors and enable less skilled actors to develop simple but effective attacks.[8] Cyber threat actors' uses of gen AI for data leakage or attacks include:

i) *Social engineering techniques*: Threat actors can use LLMs to conduct more targeted phishing attacks, business email compromise, deepfakes and other frauds. For instance, gen AI allows threat actors to misrepresent themselves more realistically as reflecting a variety of backgrounds, languages, statuses and genders.

ii) *Malware/code generation*: Gen AI can help create new malware codes or more complex variants of existing ones which can more effectively evade an automated signature-based detection system.

iii) *Disinformation*: Gen AI can increase a targeted attack's efficiency by conducting disinformation campaigns using more realistic human language characteristics and personalities.

7 A distinctive feature of gen AI models is their ability to efficiently extract information from both structured and unstructured data sources.

8 Applications such as FraudGPT and WormGPT have augmented existing cyber threats and created new dimensions of cyber risk accessible to a broader range of actors (Falade (2023)).

Furthermore, the adoption of AI tools for internal operations and cyber defence generates the risk of attacks against AI systems themselves (Zhu et al (2024)). These risks include:

i) *Data/model poisoning*: This is the process of corrupting the training data of the gen AI internal model to impair the training process or gain a desired output (Improta (2024)).
ii) *Data leakage during inference*: Threat actors can gain access to confidential data through model inversion and programmatic query of the model during the inference phase.
iii) *Vulnerability discovery*: Threat actors can use AI-based tools that are typically used for cyber defence to discover vulnerabilities and identify weaknesses in an institution's network.

Finally, gen AI system dependency on data may amplify existing challenges regarding data security and privacy, including information related to third parties. Critical financial, legal and security aspects include assessing whether training data are proprietary, how data are handled, gathered and prepared, and the quality of training data.

Overall, the rise of gen AI stands to amplify existing cyber risks and create new challenges. But this innovative technology can also be used on the defence side to strengthen cyber security and help employ more proactive fraud prevention strategies. These strategies encompass various techniques such as detection, deception and adversarial attack. As we discuss below, adequate AI programming ability of IT staff and data availability for training and testing are instrumental to reaping the benefits of AI for cyber security.

3. Gen AI in central banking

We now present the results from the survey conducted among the members of the GCRG in January 2024. We start by assessing the current status of gen AI adoption in central banks. Over two thirds (71%) of respondents are already using gen AI, and 26% have plans to incorporate such tools into their operations within the next one to two years.[9] The adoption rate could therefore approach 100% in the near term.

However, central banks are still exploring how best to integrate this technology. Specifically, when asked whether they have a concrete strategy for adopting and integrating gen AI, only 19% stated that they do, whereas 23% answered that they do not have any such strategy (Graph 1.A). A significant majority (55%) indicated that their strategy is still "in development".

Such a cautious approach can be partly attributed to the prevailing uncertainty about the correct use of gen AI. At the current stage, there are not only substantial perceived opportunities but also risks and challenges that require an adequate understanding before taking action. For a preliminary assessment of perceived

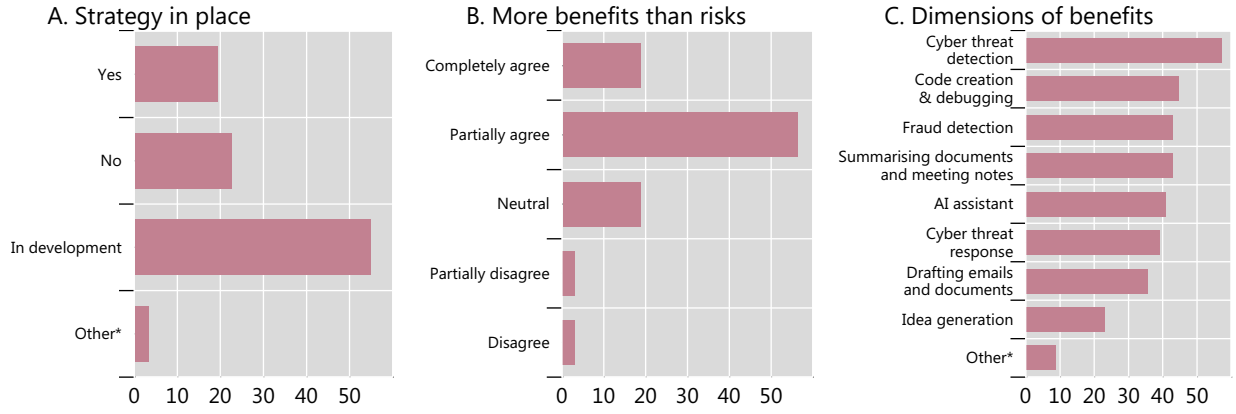
[9] These statistics are based on the answer to the following question: "Are you currently using AI systems in your organisation?"

advantages and disadvantages, the survey asked central bank experts to evaluate if the use of AI will provide more benefits than risks to their organisations. The replies, reported in Graph 1.B, are somewhat heterogeneous: 19% completely agree with the statement, 56% partially agree, 19% are neutral and 6% either partially or completely disagree. These findings suggest a general inclination among central banks to recognise the (net) advantages of gen AI.

Graph 1.C shows the dimensions along which central banks perceive benefits from gen AI. "Cyber threat detection" is the most frequently chosen dimension, selected by 57% of respondents. Additionally, central banks believe that "Code creation and debugging", "Fraud detection" and "Cyber threat response" can also gain from gen AI implementation, albeit to a lesser degree. These responses highlight that cyber attacks are a serious concern for central banks. Other dimensions of benefits include activities usually included in employees' daily work, such as "Summarising documents and meeting notes" (43% of respondents) and "Drafting emails and documents" (36% of respondents).

Adoption of gen AI in central banking

As a percentage of respondents — Graph 1



Panel A reports the share of respondents that selected each answer to the question "Does your organisation have a strategy in place regarding the evaluation and adoption of AI?". Other* includes: "Rather simple guidelines that will be amended from time to time". Panel B reports the share of responses to the question "Do you agree that the use of AI can provide more benefits than risks to your organisation?". Panel C reports the share of respondents that selected each option when asked "Where do you think your organisation could benefit from the use of AI?"; respondents could choose multiple options. Other* includes: "Productivity-suite efficiencies, creating outlines/drafts, ATIP/search intelligence"; "Decision support system, physical security"; "Chatbot mode to help a user finding information"; "Financial oversight use cases as well like regtech"; "Early warning of financial stability tasks, forecasting of key economic indicators, and work efficiency enhancement".

Source: Authors' calculations.

4. Opportunities, risks and challenges for cyber security

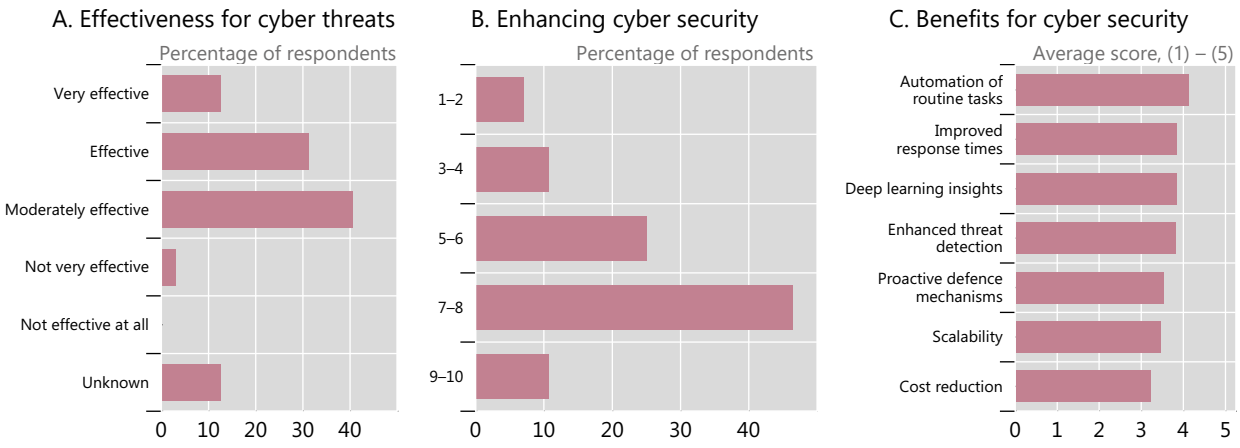
A key challenge for central banks is to set up an IT infrastructure that can effectively leverage the benefits of gen AI while addressing current and foreseen cyber risks. This section offers a more detailed analysis of the opportunities (Section 4.1) as well as the risks and challenges (Section 4.2) associated with the widespread adoption of gen AI, as perceived by surveyed leaders within central bank IT cyber security units.

4.1 Opportunities

Traditional machine learning tools for cyber risk management have been employed by central banks for some time. These tools include systems for detecting and responding to cyber threats, securing transactions and monitoring the integrity of payment systems.

Gen AI could enhance cyber security abilities beyond the capabilities of more traditional methods and contribute to enhancing cyber security more broadly. When explicitly asked to evaluate this aspect, 44% of respondents perceive gen AI to be either very effective or effective, whereas 41% consider it moderately effective (Graph 2.A). Only 3% believe the new technology to be not very effective, with 13% abstaining from providing an evaluation. In line with the previous question, the survey invited central banks to assess the impact of gen AI on enhancing cyber security measures (Graph 2.B). Using a scale from 1 (low enhancement) to 10 (high enhancement), one quarter of the respondents rated this between 5 and 6, over 45% between 7 and 8 and 11% rated it 9 to 10, while 7% chose a value lower than or equal to 2. The perceived capability of gen AI to boost cyber security is thus broadly positive, but its overall impact remains somewhat uncertain.

Effectiveness of gen AI for defence against cyber threats — Graph 2



Panel A reports the share of respondents that selected each answer to the question "How effective do you believe AI is in identifying and responding to cyber threats compared to traditional methods?". Panel B reports respondents' ratings on a scale of 1 (very low) to 10 (very high) in response to the question "Overall, how would you rate the impact of AI on enhancing cyber security measures on a scale of 1–10 in your opinion?"; answers have been grouped in five bins. Panel C reports the average score respondents gave to each option when asked to "Rate the level of significance of the following benefits of AI in cyber security"; the score scale of each option is from 1 (lowest) to 5 (highest).

Source: Authors' calculations.

The mixed assessment of the potential benefits of adopting AI in cyber security could stem from the different operational areas where experts anticipate the most significant gains. Graph 2.C illustrates the average scores assigned to various facets of cyber security that could benefit from AI, with ratings on a scale from 1 (very low benefits) to 5 (very high benefits). "Automation of routine tasks" received the highest average score, indicating that a key advantage of using gen AI tools lies in their ability to replace labour-intensive tasks typically performed by humans. The dimensions receiving the next-highest ratings are "Improved response times", "Deep learning insights" (referring to the deep learning aspect of AI capable of offering insights into

cyber threats by analysing data patterns beyond human capabilities) and "Enhanced threat detection". This underscores the belief that gen AI will have a positive impact on both the detection of cyber threats and the response to attacks. More broadly, the patterns in Graph 2 suggest that AI is likely to broaden the scope and size of cyber security units at central banks through a combined process of automating routine tasks and investing in new soft skills.

4.2 Risks and challenges

The increasing sophistication and frequency of cyber threats, along with the advent of gen AI tools, introduce novel risks and challenges to the cyber security frameworks of regulatory and supervisory authorities.

The survey asked respondents to rate their concerns regarding various sources of vulnerabilities that gen AI systems can introduce into cyber security defences. For all vulnerabilities reported in Graph 3.A, respondents were asked to assign a score from 1 (lowest level of concern) to 5 (highest level of concern). Respondents are, on average, most concerned about "Social engineering" and "Unauthorised data disclosure", which are not unrelated risks. Gen AI tools are enabling increasingly sophisticated cyber attacks, notably through their ability to replicate voices or images and create deepfakes, aiming ultimately to extract information or infiltrate networks.[10] Implications of a social engineering attack include unauthorised access to banks' internal networks (with potential blocking of the system and demands for ransom payments) as well as unintentional disclosure of sensitive data. Such attacks can undermine trust in central banks and, in the extreme, jeopardise financial stability. These concerns extend to all employees at central banks, not only those in IT units, underscoring the need for a comprehensive policy that educates the entire staff on the implications of adopting gen AI tools. Similarly, the next highest-rated concern is "Privacy issues" – the risk associated with the potential exposure of sensitive data – followed by "Black box algorithms" and "Automated propagation". These concerns may reflect fears of a "loss of control" following the adoption of gen AI tools by cyber security operators, as decisions made by autonomous AI models can be hard to understand and explain. Overall, when asked to evaluate the risk associated with implementing gen AI for cyber security on a scale of 1 (very low) to 10 (very high), 96% of respondents answered with 5 or higher.[11]

The survey also asked respondents to rate different challenges arising from the use of gen AI for cyber security on a scale from 1 (lowest) to 5 (highest). Graph 3.B reports how central banks assess the challenges that are typically encountered in the integration (or planned integration) of gen AI into existing cyber security systems. For most options, central banks see substantial challenges. Notably, the "Skill gap" (the shortfall of individuals proficient in both AI programming and cyber security) receives the highest average score. The next section will revisit this challenge in more detail.

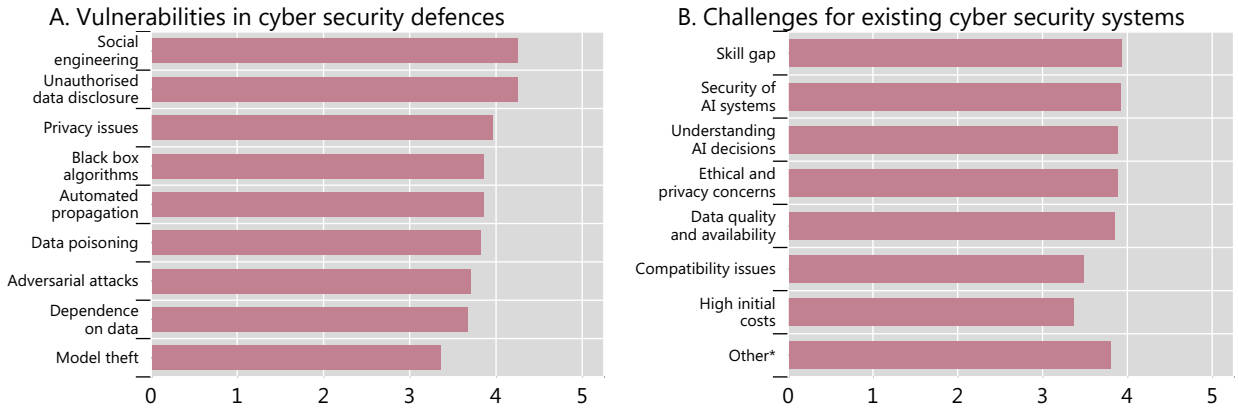
[10] A strong concern for central banks is the risk that employees might become targets of cyber attacks triggered by AI, such as highly personalised phishing emails or other forms of social engineering frauds.

[11] These statistics are based on the answer to the following question: "Based on a scale from 1 to 10, overall, how would you rate the risk associated with implementing AI in cyber security?".

Other challenges include issues related to data management and transparency. This includes "Security of AI systems" (ensuring AI systems are safeguarded against vulnerabilities and misuse by attackers),[12] "Understanding AI decisions" (the difficulty in interpreting and understanding the decision-making processes of AI, especially with complex algorithms), "Ethical and privacy concerns" (addressing the ethical implications and privacy issues arising from AI use in surveillance and other security functions) and "Data quality and availability" (challenges related to the availability of high-quality, relevant data needed to train AI models effectively).

Risks and challenges posed by AI adoption for cyber security

Average score, 1 (very low)–5 (very high) Graph 3



Panel A reports the average score given to each option when respondents were asked to "Rate your concern regarding the following vulnerabilities AI systems themselves may introduce into cyber security defences"; the scale of each option is from 1 (very low) to 5 (very high). Panel B reports the average score given to each option when respondents were asked to "Rate the level of challenge posed by the following when integrating or planning to integrate AI into existing cyber security systems"; the scale of each option is from 1 (very low) to 5 (very high). Other* includes: "Data access, data hosting and data transfer (out of the country) with regard to regulatory constraints"; "Educating users on the limits and risks of this current generation of AI, as it can speak authoritatively even when completely wrong"; and "Being driven by suppliers with their best interest at heart. Lack of greater good cooperation".

Source: Authors' calculations.

5. IT investments and human capital

Most central banks have significantly increased their annual budget for investment in cyber security since 2020 (Doerr et al (2022)). The rise of gen AI reinforces this trend and calls for an urgent update in skills, through both training initiatives for existing staff and hiring of new employees.

When discussing the relation between the level of human capital and the integration of gen AI with cyber security, two important dimensions of analysis arise.

[12] A specific security threat that is becoming increasingly popular is "model poisoning", where attackers deliberately introduce harmful data or manipulate the training process of an AI system to compromise its integrity or functionality. This attack aims to alter the model's predictions or behaviour in a way that serves the attacker's purposes (see Hitaj et al (2022, 2024)).

The first pertains to all employees at the central bank, focusing specifically on the adoption of gen AI tools. This adoption may be hindered by the existing workforce's insufficient technological skills or could even be prevented by internal practices due to concerns related to data protection and privacy. The second relates more directly to the human capital within the IT divisions of central banks, specifically concerning cyber security.

Concerning the first dimension, the survey inquired whether central banks have enabled or plan to enable their staff to access cloud-based gen AI applications (eg ChatGPT). Most respondents replied affirmatively, albeit with certain restrictions (eg prohibiting the submission of corporate information). Among respondents, 13% reported not currently having this capability but were planning to implement it soon, whereas 9% indicated no plans to enable access. Unrestricted access to cloud-based applications for staff has been allowed by only 3% of respondents.[13]

Generally speaking, cyber security experts are concerned about the preparedness and capability of current staff to effectively integrate and utilise AI systems. When asked about this, 40% of respondents expressed high or extreme concern, 35% moderate concern, 15% slight concern, and one respondent reported being unconcerned.[14] An implication is that introducing internal practices and common regulatory policies might be necessary for a safe and widespread adoption of gen AI tools in central banks.

A second distinct issue concerns the IT investments and management of human capital employed in the cyber security units of central banks. Graph 4.A shows that the most frequently mentioned benefit of gen AI is "Increased efficiency", highlighting the ability of cyber security staff to utilise tools that facilitate quicker decision-making and responses. Another key advantage is "Reduced workload", attributed to the automation of numerous processes that lessen the burden on cyber security personnel. Some central banks have reported "No significant change", indicating that gen AI has been integrated without significant shifts in the allocation of human resources. Lower ratings were given to "Skillset upgradation" (the necessity to enhance skills within the cyber security workforce to effectively manage and interact with AI tools) and "Reallocated for strategic tasks" (with AI assuming routine tasks, thus freeing human resources to concentrate on more strategic cyber security efforts). Overall, while gen AI stands to provide benefits in terms of routine tasks, cyber security units are simultaneously focusing on acquiring new skills for their existing staff and reallocating them towards more strategic cyber security initiatives.

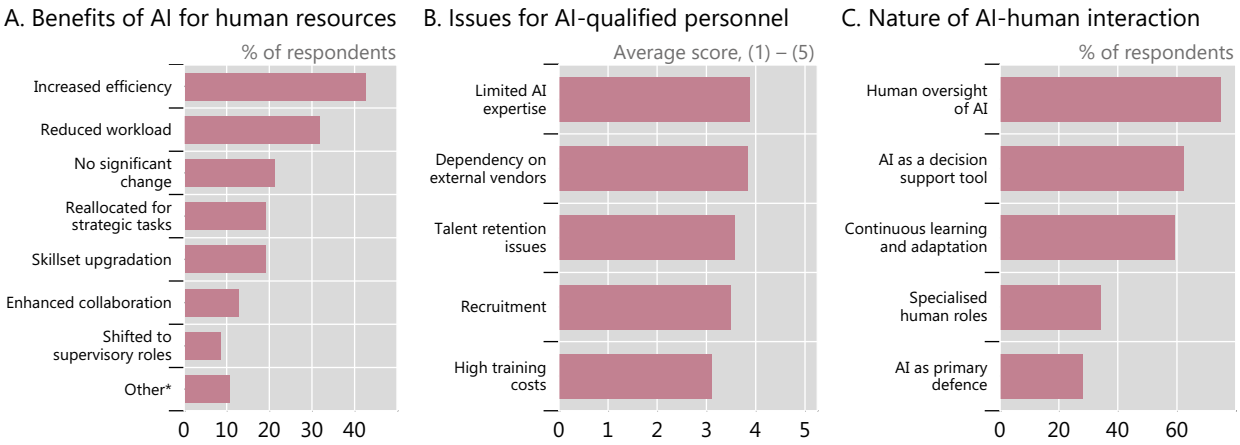
Concerns persist regarding the scarcity of AI-qualified personnel. For a range of listed concerns, respondents were asked to assign a score from 1 to 5, with 1 being the lowest and 5 the highest (Graph 4.B). "Limited AI expertise" emerged as the highest-rated concern, underscoring the significance of talent retention and recruitment in personnel management. The second-highest concern is "Dependency on external vendors", indicating a substantial reliance on external providers for AI

[13] These statistics are based on the answer to the following question: "Have you enabled/are you planning to enable staff access to cloud-based generative AI applications, such as ChatGPT, Claude, Bard, Copilot or Midjourney?"

[14] These statistics are based on the answer to the following question: "What is your level of concern regarding the preparedness and capability of your current staff in effectively onboarding and operationalising AI systems within your organisation?"

solutions, which may also be a function of inadequate internal personnel. Dependence on cloud services has already been recognised as crucial for cyber risk management in central banks, particularly in advanced economies (Doerr et al (2022)). Gen AI gives these concerns new impetus (Araujo et al (2024)).

AI, cyber security and human capital of central banks — Graph 4



Panel A reports the share of respondents choosing each option when asked "How may AI impact or how has it already impacted the allocation of human resources in cyber security tasks?". Other* includes: "We think it is still early for us to see real improvement in terms of workload optimisation through the use of AI-based tools"; "In many cases, the impact on HR is still unknown"; "We have no idea yet how this will transform the way we work in cyber security"; "No impact yet. Use of AI is limited as pilot testing and access or interface to production data are not yet allowed". Panel B reports the average score on a scale of 1 (very low) to 5 (very high) that respondents gave to each option when asked to "Rate the following concerns regarding the limited availability of AI-qualified personnel". Panel C reports the share of respondents choosing each option for the question "How do you envision the interaction between AI systems and human cyber security experts evolving?".

Source: Authors' calculations.

Finally, the survey explores central banks' views on the collaboration between AI systems and human cyber security experts. The majority of respondents recognise AI as an opportunity, albeit one that requires human supervision (Graph 4.C). Gen AI is predominantly seen as a tool to support human experts, enhancing their productivity rather than serving as the primary, autonomous defence mechanism. Experts in cyber security units also agree on the fact that humans will continuously train and update AI systems, while also utilising their continuous learning. Overall, these patterns suggest a complementary relationship between human and AI capabilities.

6. Future perspectives and regulatory insights

So far, we have analysed the issues regarding the adoption of gen AI in cyber security practices of central banks, mostly providing a snapshot of the current situation. The survey also asked about central banks' assessment of future aspects of cyber risk management and their perspective on actions that can be put forward to adequately incorporate gen AI into their operations.

In recent years, central banks have significantly invested in several areas related to cyber security. These include developing incident response plans in case of cyber

attacks (for example with internal exercises to simulate attacks), providing risk management frameworks for cyber security (eg by conducting cyber stress tests) and collecting information on cyber attacks on financial institutions. More generally, there has been a shift from a compliance-based focus to a risk management and resilience focus.

An open question is whether gen AI will change overall strategies and approaches to cyber security in the coming years. Survey results indicate that most respondents do not foresee major changes. Beyond expected changes due to the adoption of the new technology, the evolving landscape confirms the current trend towards a risk management and resilience approach to cyber security. Most central banks anticipate that gen AI systems will lead to a shift towards proactivity – that is, shifting from a reactive to a proactive stance in predicting and neutralising threats before they manifest. AI is also expected to provide a more customised defence of networks and systems, based on user behaviour and company profiles, alongside a dynamic risk assessment for detecting and defending against new threats. All in all, gen AI tools are predicted to support risk management functions and enhance existing protocols, without altering the recent shift towards an improvement of cyber resilience – that is, the capacity of central banks to foresee, adapt to and swiftly recover from cyber incidents while continuing their essential operations.[15]

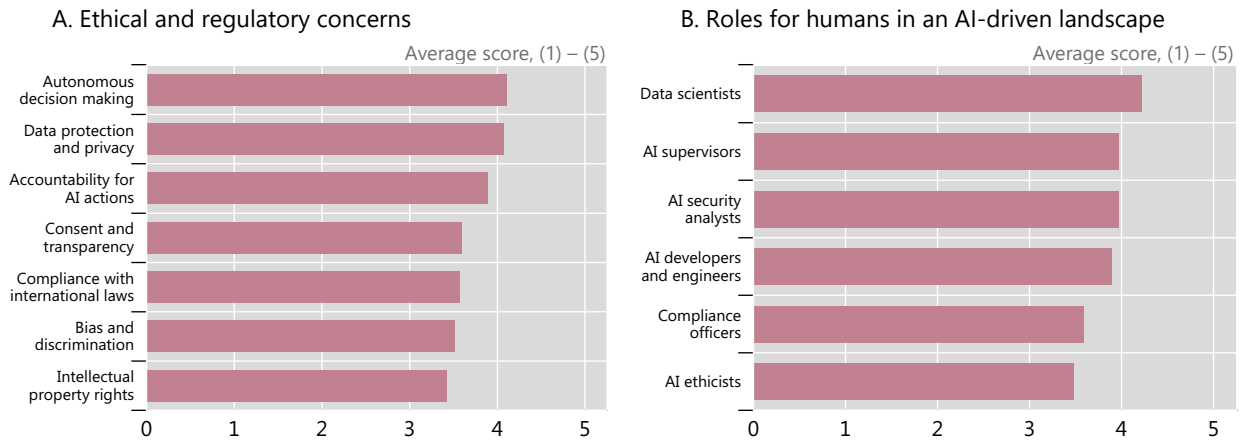
The survey also asked participants to rate various ethical and regulatory concerns should gen AI become increasingly integrated with cyber security. The rating scale ranges from 1 to 5, with 1 being the lowest and 5 the highest concern. Graph 5.A shows that the primary concerns identified are "Autonomous decision-making", which involves setting limits on decisions AI can make independently, and "Data protection and privacy", emphasising the importance of maintaining stringent data protection and privacy norms amidst AI's enhanced data analysis capabilities. Another significant issue highlighted is "Accountability for AI actions" (establishing guidelines to define responsibility for AI-driven decisions, particularly when they result in adverse outcomes). "Compliance with international laws" and "Consent and transparency" are considered relatively less urgent issues at the moment, probably due to the ongoing adaptation and refinement of regulations to keep pace with technological progress.

What roles will become increasingly critical for human workers as the use of gen AI expands? The survey asked participants to rate different aspects on a scale from 1 (lowest) to 5 (highest). "Data scientists", with their expertise in understanding and interpreting data integral to AI systems as well as refining AI learning processes, are expected to play a major role (Graph 5.B). This category is followed closely by "AI supervisors" (who ensure that AI operations adhere to organisational objectives and ethical standards), "AI security analysts" (specialists in safeguarding AI systems against breaches or manipulation by malicious entities) and "AI developers and engineers" (the technical experts responsible for creating, sustaining and enhancing AI algorithms and systems designed for cyber security). These results confirm the importance of human expertise for the correct adoption and use of gen AI tools for cyber security. Emphasis is also placed on professionals who ensure adherence to

[15] These conclusions are based on the answers to the following question: "How will AI change the overall strategies and approaches to cyber security in the coming years?", which highlight different dimensions of cyber security strategies.

ethical principles and societal norms, as well as compliance officers who verify that AI cyber security practices meet legal, regulatory and industry standards.

The future landscape of AI and cyber security — Graph 5



Panel A reports the average score that respondents gave to each option when asked to "Rate the following ethical and regulatory concerns as AI becomes more prevalent in cyber security"; the scale of each option is from 1 (very low) to 5 (very high). Panel B reports the average score that respondents gave to each option when asked to "Rate the following roles that may become more crucial for humans in an AI-driven cyber security landscape"; the scale of each option is from 1 (very low importance) to 5 (very high importance).

Source: Authors' calculations.

7. Conclusion

Cyber attacks are becoming increasingly frequent and evolving in complexity and sophistication. At the same time, there are significant shifts in technology generated by the rapid developments in gen AI systems.

By conducting a tailored survey of central bank cyber security leaders from the CRCC-administered GCRG forum in January 2024, we investigate the status of adoption of gen AI tools, identify the perceived benefits and risks associated with their use for cyber security, and highlight perceived challenges and future perspectives from the viewpoint of central banks. While AI regulation is not yet fully developed (Aldasoro et al (2024b)), there is a strong consensus on the adoption of common rules for the use of AI for cyber security, and a recognition that new forms of cooperation at the central bank level are needed. Such collaborative efforts should address the establishment of new data protection standards to ensure the responsible implementation of gen AI and, crucially, tackle the issue of the "Skill gap" among human personnel.

The BIS supports the cyber security efforts of central banks and global cooperation through the CRCC. Established in 2019, the CRCC plays a pivotal role in the future integration of gen AI for cyber security. A structured approach to knowledge-sharing, collaboration and training of human capital is of first-order importance to face future challenges. One of the key projects is the GCRG forum, which includes CISOs from BIS member central banks, representing cyber security

leadership in the global central bank community. This group is instrumental in addressing the challenges presented by the adoption of AI technologies. Other CRCC initiatives include a global cyber resilience collaboration platform with over 300 active cyber security professionals from the central bank community. This platform and community are poised to become a central knowledge-sharing and collaboration forum on the topic of AI challenges and adoption.

The CRCC also leads the Cyber Resilience Assessments project. Its purpose is to provide central banks with a common framework to assess their cyber resilience posture and improve their cyber resilience practices in the delivery of critical business services. The CRCC has performed cyber resilience assessments across various BIS member central banks and delivered a global benchmark for the central bank community. Central banks can now compare their cyber resilience posture to the benchmark and make informed investment decisions regarding cyber security. Moreover, the CRCC also holds various community events such as annual cyber security seminars and cyber range exercises. These events assist in keeping the global central bank cyber security community engaged with emerging cyber security issues and threats (such as AI adoption), thereby ensuring operational readiness.

Cooperation and information-sharing are key to collectively reducing cyber risk and preventing and containing major cyber incidents. The importance of cooperation is set to further increase with the adoption and development of gen AI systems.[16]

[16] Under the stimulus of standard-setting bodies such as the Financial Stability Board or G7, international work and cooperation on cyber security is ongoing. See Basel Committee on Banking Supervision (2018, 2021), Financial Stability Board (2020) and G7 (2016).

References

Aldasoro, I, S Doerr, L Gambacorta and D Rees (2024a): "The impact of artificial intelligence on output and inflation", *BIS Working Papers,* no 1179, April.

Aldasoro, I, J Frost, L Gambacorta, T Leach and D Whyte (2020): "Cyber risk in the financial sector", *SUERF Policy Notes*, no 206, November.

Aldasoro, I, L Gambacorta, P Giudici and T Leach (2022): "The drivers of cyber risk", *Journal of Financial Stability*, vol 60, June.

Aldasoro, I, L Gambacorta, A Korinek, V Shreeti and M Stein (2024b): "Intelligent financial system: how AI is transforming finance", *BIS Working Papers*, forthcoming.

Araujo, D, G Bruno, J Marcucci, R Schmidt and B Tissot (2022): "Machine learning in central banking", *IFC Bulletins*, no 57, November.

Araujo, D, S Doerr, L Gambacorta and B Tissot (2024): "Artificial intelligence in central banking", *BIS Bulletins*, no 84, January.

Basel Committee on Banking Supervision (2018): *Cyber-resilience: range of practices*, December.

——— (2021): "Newsletter on cyber security", September.

Boissay, F, G Cornelli, S Doerr and J Frost (2022): "Blockchain scalability and the fragmentation of crypto", *BIS Bulletins*, no 56, June.

Brynjolfsson, E, D Li and L Raymond (2023): "Generative AI at work", *NBER Working Papers*, no 31161.

Doerr, S, L Gambacorta and J Serena Garralda (2021): "Big data and machine learning in central banking", *BIS Working Papers*, no 930, March.

Doerr, S, L Gambacorta, T Leach, B Legros and D Whyte (2022): "Cyber risk in central banking", *BIS Working Papers*, no 1039, September.

Enterprise Strategy Group and Information Systems Security Association (2020): *The life and times of cybersecurity professionals 2020*, July.

Falade, P (2023) "Decoding the threat landscape: ChatGPT, FraudGPT, and WormGPT in social engineering attacks", arXiv:2310.05595.

Felten, E, M Raj and R Seamans (2021): "Occupational, industry, and geographic exposure to artificial intelligence: a novel dataset and its potential uses", *Strategic Management Journal*, vol 42, no 12, pp 2195–217.

Financial Stability Board (2020): *Effective practices for cyber incident response and recovery: final report*, October.

G7 (2016): *G7 fundamental elements of cybersecurity for the financial sector*, October.

Hitaj, D, G Pagnotta, F De Gaspari, D Ruko, B Hitaj, L Mancini and F Perez-Cruz (2024): "Do you trust your model? Emerging malware threats in the deep learning ecosystem", arXiv:2403.03593v1.

Hitaj, D, G Pagnotta, B Hitaj, L Mancini and F Perez-Cruz (2022): "MaleficNet: hiding malware into deep neural networks using spread-spectrum channel coding," in V Atluri, R Di Pietro, C Jensen and W Meng (eds), *Computer Security – ESORICS 2022, Lecture Notes in Computer Science*, vol 13556, Springer, Cham.

Improta, C (2024): "Poisoning programs by un-repairing code: security concerns of AI-generated code", arXiv:2403.06675v1.

Kashyap, A and A Wetherilt (2019): "Some principles for regulating cyber risk", *AEA Papers and Proceedings*, vol 109, May, pp 482–7.

McKinsey (2023): *The economic potential of generative AI: the next productivity frontier*, McKinsey Digital report, June.

Neupane, S, I Fernandez, S Mittal and S Rahimi (2023): "Impacts and risk of generative AI technology on cyber defense", 10.48550/arXiv.2306.13033.

Noy, S and W Zhang (2023): "Experimental evidence on the productivity effects of generative artificial intelligence", *Science*, vol 381, no 6654, July, pp 187–92.

Peng, S, W Swiatek, A Gao, P Cullivan and H Chang (2024): "AI revolution on chat bot: evidence from a randomized controlled experiment", mimeo.

US Department of the Treasury (2024): *Managing artificial intelligence-specific cybersecurity risks in the financial services sector*, March.

Zhu, B, N Mu, J Jiao and D Wagner (2024) "Generative AI security: challenges and countermeasures", arXiv:2402.12617v1.

Previous volumes in this series

No	Title	Issue date
BIS Papers No 144	The economic implications of services in the metaverse	February 2024
BIS Papers No 143	Central banking in the Americas: Lessons from two decades	November 2023
BIS Papers No 142	Inflation and labour markets	November 2023
BIS Papers No 141	Will the real stablecoin please stand up?	October 2023
BIS Papers No 140	Central banks, macro-financial stability and the future of the financial system	October 2023
BIS Papers No 139	Digital safety nets: a roadmap	September 2023
BIS Papers No 138	Financial stability risks from cryptoassets in emerging market economies	August 2023
BIS Papers No 137	Building an integrated surveillance framework for highly leveraged NBFIs – lessons from the HKMA	July 2023
BIS Papers No 136	Making headway – Results of the 2022 BIS survey on central bank digital currencies and crypto	July 2023
BIS Papers No 135	The energy transition and its macroeconomic effects	May 2023
BIS Papers No 134	Global tightening, banking stress and market resilience in EMEs	April 2023
BIS Papers No 133	The two-regime view of inflation	March 2023
BIS Papers No 132	Information governance in sustainable finance	December 2022
BIS Papers No 131	Central banking after the pandemic: challenges ahead	December 2022
BIS Papers No 130	Pricing of climate risks in financial markets: a summary of the literature	December 2022
BIS Papers No 129	The role of non-bank financial institutions in cross-border spillovers	December 2022
BIS Papers No 128	Central bank digital currencies in Africa	November 2022
BIS Papers No 127	Historical monetary and financial statistics for policymakers: towards a unified framework	September 2022

All volumes are available on the BIS website (www.bis.org).