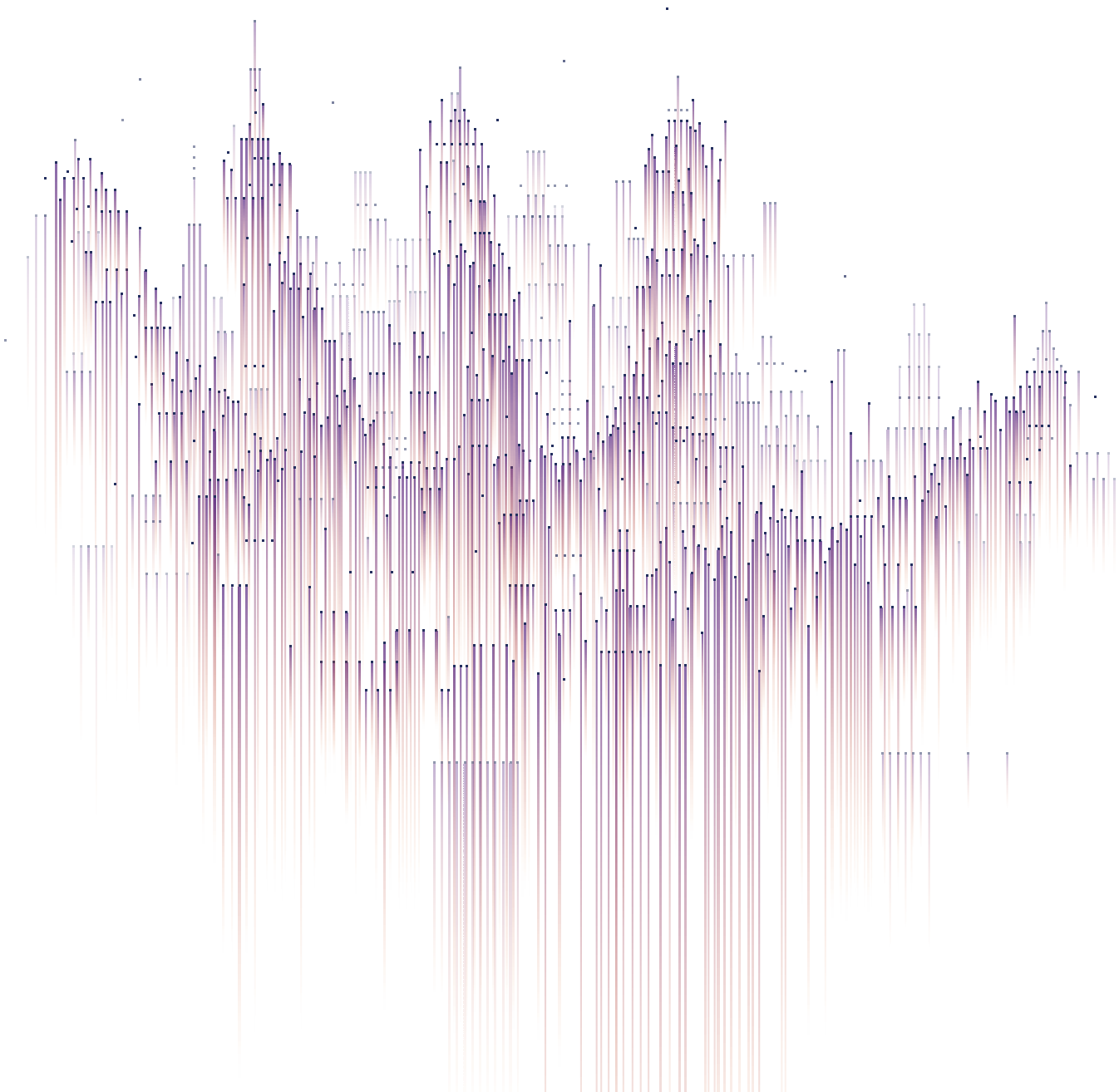
GCHQ



Pioneering a New National Security

The Ethics of Artificial Intelligence

Contents

Foreword by Director GCHQ

Innovation, technology and data is embedded deep in the organisational DNA of GCHQ. Everyone who meets our analysts, engineers or apprentices is struck by their passion for keeping the UK safe and their willingness to champion new ideas and approaches.

Soon after being appointed as Director of GCHQ, I asked to meet with some of the best data scientists from across the agency. The examples they showed me of their existing work were eye-opening; their excitement about the potential of Artificial Intelligence (AI) to transform our future operations was palpable. Their forebears at Bletchley Park would have been proud.

Four years on, and AI is now present in every aspect of British life. It enables our telecommunications systems, our smartphones, our banks, our National Health Service. The UK's global leadership in AI and data science is a major part of what has made the UK a thriving cyber power, and AI stands to add billions to the British economy.

But the situation has not stood still. The COVID-19 pandemic has shown us how quickly things change. Technology has allowed us to respond. However, it's also clear that the UK's adversaries have not stood still. The nation's security, prosperity and way of life faces new threats from hostile states, terrorists and serious criminals, often enabled by the global internet. An ever-growing number of those threats are to the UK's digital homeland – the vital infrastructure and online services that underpin every part of modern life.

At GCHQ, we believe that AI capabilities will be at the heart of our future ability to protect the UK. They will enable analysts to manage the ever-increasing volume and complexity of data, improving the quality and speed of their decision-making. Keeping the UK's citizens safe and prosperous in a digital age will increasingly depend on the success of these systems.

Rapid technological change always generates deep, challenging questions, and the development of AI is no exception.

Philosophers and data scientists have been grappling with the implications of AI for ethics: how do you ensure that fairness and accountability is embedded in these new systems, for example? How do you prevent AI systems replicating existing power imbalances and discrimination in society? Debate on these and many other questions is still in the early phases. GCHQ is sponsoring work through the Alan Turing Institute and other civil society institutions to help provide practical answers.

Similar discussions are occurring in international institutions from New York to Geneva. How should the traditional international rules-based system respond to AI and other emerging technologies? How can governments and citizens build institutions capable of engaging with this digital age? The UK's Presidency of the G7 during 2021 will see an ambitious sequence of events to explore and develop these themes, in which GCHQ will be closely involved.

These are major issues of international importance and this paper can only hope to begin a conversation on the way ahead. It is clear that making progress is critical for GCHQ, as it is for every other area of life. I hope you find this paper as interesting as I have, and look forward to continuing this debate over the coming few years.

Executive Summary

Britain today is a digital nation, leading and shaping events across a world inextricably linked through cyberspace. Now and into the future, the value of our economy, our way of life, and our global influence will be built on our advanced digital infrastructure, capabilities and knowledge.

Artificial Intelligence – a form of software that can learn to solve problems at a scale and speed impossible for humans – is increasingly essential to the way we live. It is already transforming sectors as diverse as healthcare, telecommunications, and manufacturing. AI software informs our satnavs, guides our internet searches, and protects us every time we make an electronic purchase, or open an app on our smartphone.

In the century since it was founded, GCHQ has been at the forefront of innovation in national security. Generations of brilliant analysts, with their diverse mix of minds, have used their technical ingenuity, cutting-edge technology and wide-ranging partnerships to identify, analyse and disrupt threats to our nation.

Today, as technological change continues to accelerate, we are pioneering new approaches to understanding the complex and interconnected world around us. We have long championed the responsible use of data science, and believe that AI will be at the heart of our organisation's future.

Thinking about AI encourages us to think about ourselves, and what it means to be human: our preferred way of life, our guiding values and our common beliefs. The field of AI ethics has emerged over the last decade to help organisations turn these ethical principles into practical guidance for software developers – helping to embed our core values within our computers and software.

We won't pretend that there are not challenges ahead of us. In using AI we will strive to minimise and where possible eliminate biases, whether around gender, race, class or religion. We know that individuals pioneering this technology are shaped by their own personal experiences and backgrounds. Acknowledging this is only the first step – we must go further and draw on a diverse mix of minds to develop, apply and govern our use of AI. Left unmanaged, our use of AI incorporates and reflects the beliefs and assumptions of its creators – AI systems are no better or no worse than the human beings that create them.

Our society is learning and growing: the Alan Turing Institute and similar bodies are helping to show us how we might build and use AI in a more ethical, responsible manner. GCHQ is committed to creating and using AI in a way that supports fairness, empowerment, transparency and accountability – and to protecting the nation from AI-enabled security threats pursued by our adversaries. We believe that, by working together with our partners across Britain and beyond, we can deliver this vision.

This paper describes the digital Britain of today, and our values-led approach for the spaces where people, information and technology meet. It lays out GCHQ's AI and Data Ethics Framework, and how we intend to use AI in our operations. It forms part of our commitment to inclusion, debate and openness. The paper is the first step of a much longer journey: we'd like you to join us on it.

Key Insights

An increasing use of AI will be fundamental to GCHQ's mission of keeping the nation safe. AI will be vital to our ability to deal with the ever-increasing volume and complexity of data, and to develop the capabilities needed to defend against AI-enabled threats by malicious actors.

We operate within an internationally acclaimed legal and regulatory framework, which balances the preservation of important individual rights and liberties with protection from significant threats to our way of life. Independent oversight ensures that the way in which we exercise our powers, including through the application of AI, is done in accordance with the law.

We take our Privacy and Human Rights obligations very seriously. To ensure the impact on privacy is properly considered in every circumstance, we undertake an assessment to determine the necessity and proportionality of any intrusion into privacy, both when considering the use of operational data to train and test AI software, and when applying the software to the analysis of operational datasets. These assessments are made available for audit by the Investigatory Powers Commissioner's Office (IPCO).

We are developing a comprehensive governance system to manage AI and data ethics, drawing on best practice and consultation with a wide range of external stakeholders. It will set out the standards our developers will be expected to meet and practical guidance to help them achieve this, together with a supporting educational programme.

The development and deployment of AI for national security presents some unique ethical challenges. Specialist expertise and highly developed processes will be needed to minimise bias and discrimination, and ensure the production of accurate results.

In GCHQ, AI will be about enabling humans to make better decisions. Our efforts will focus on developing "augmented intelligence"(AuI) systems, utilising AI to collate information from relevant sources and highlight significant data for review by our analysts; supporting the decision-making process rather than determining it.

To ensure our use of AI remains ethical, the governance system will need to grow as the technology evolves and we find new ways to utilise it. We will continue to engage and learn to ensure our guidance reflects current thinking.

Artificial Intelligence at GCHQ

Introduction

Picture a cold and rather drafty morning at Bletchley Park, on 18 January 1944. Some of the finest minds in modern mathematics and engineering have been working together for several years in support of the Allied war effort.

It has become clear, however, that the electromechanical computers used by the teams can no longer keep up with the Nazi communication systems. For six months, telecoms engineer Tommy Flowers has been working in strict secrecy on a replacement. That morning, the mathematician Max Newman records a simple statement: "Colossus arrives today".

All of GCHQ's modern supercomputers and racks of advanced Cloud processors stem from the computer that was delivered to Bletchley Park that day. Colossus weighed several tons and depended on several thousand vacuum tubes to function, but it was the world's first digital computer, far faster and more effective than anything that had gone before it. It tore through the vast sets of communications data generated by the Nazi military hierarchy, quickly proving itself to be a key element in the Allied response.

Colossus was, of course, a very basic system by modern standards. Constructed laboriously by hand, it was fed information using reams of paper tape, manually glued together at each end to make an extended loop. It could not store information permanently in its memory and being unable to adapt itself to new tasks, depended on its human technicians at every step.

Its electronic descendants around the globe, all tracing their path back to the original Colossus, have infinitely more processing power and data storage. Their mathematical algorithms can consume and manipulate information on a scale that would have been unimaginable in 1944. With the help of an ever-increasing number of computer scientists, they can be trained to perform some tasks previously only possible for humans. After 70 years, AI has finally come of age.

Artificial Intelligence

There are many definitions of AI around the world, but at GCHQ we think of AI as a type of software that can learn to find complex patterns in data. As the software does this, it can provide us with new insights or forecast future trends. We are able to take these conclusions, and use them to automate or augment business processes. In some cases, we could use them to manage quite basic, highly-repetitive activities, while in others we might tackle more sophisticated but narrowly defined challenges.

Enabled by inexpensive, accessible computer processing power and the availability of large quantities of data, AI is transforming the world around us. A new generation of internet-based companies has rapidly evolved to exploit the huge value of data when analysed using AI. As Professor Dame Wendy Hall and Jérôme Pesenti suggested in 2017:

"AI offers massive gains in efficiency and performance to most or all industry sectors, from drug discovery to logistics. AI…can be integrated into existing processes, improving them, scaling them, and reducing their costs, by making or suggesting more accurate decisions through better use of information."[1]

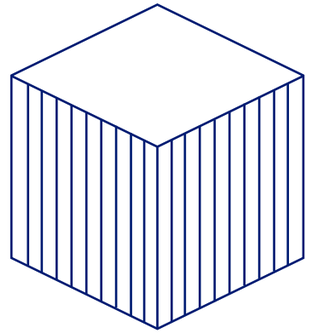
Unlike Colossus in 1944, the development of AI today is no niche, secret activity, but one which applies to all sectors and will underpin global prosperity, innovation and jobs. The opportunities it creates are changing the focus of organisational and social development. As the United Nations Education, Scientific and Cultural Organisation highlighted in 2019:[2]

"Very soon, the tools of education – the way we learn, access knowledge and train teachers – will no longer be the same. From now on, the acquisition of digital skills stands at the centre of all our education programmes. Furthermore, we must 'learn to learn' because the pace of innovation is rapidly transforming the labour market."

The UK – with its world-class universities and industry – is at the vanguard of AI development. Under our national Industrial Strategy, AI has been declared one of the UK's "Grand Challenges". The UK is committed to improving the institutions that support AI development, building a skilled workforce, and stimulating access to data. The potential prize is huge: experts suggest that investment in AI could add an additional £630 billion to the UK economy by 2035, increasing our annual growth rate by up to 3.9%.[3]

AI and GCHQ

At GCHQ, we take forward pioneering security at the cutting edge of technology. AI will be a critical issue for our national security in the 21st century. We know that AI software will be an increasingly powerful analytical tool in our mission to keep the UK safe, and that keeping the AI systems underpinning the UK's prosperity secure from cyber attack will become an ever more important part of this work.



Although many are excited by the opportunities that AI presents for the UK's economic growth and security, others are reflecting on the potential social and ethical risks that surround these changes. Leading figures in civil society and academia are researching the potential for AI software to cause harm to society, whether as a result of poor design or even deliberate misuse. As a security and intelligence agency, GCHQ is also aware that the UK's adversaries are actively considering how to use AI against our national interests.

We believe that the UK needs increased dialogue and debate around the use and protection of AI, backed by expert technological, operational and ethical insights, and supported by wide public engagement. As Jeremy Fleming, the Director of GCHQ, has argued:[4]

"We need honest, mature conversations about the impact that new technologies could have on society. This needs to happen while systems are being developed, not afterwards. And in doing so we must ensure that we protect our [citizens'] right to privacy and maximise the tremendous upsides inherent in the digital revolution."

GCHQ is publishing this paper as part of that national dialogue, to explore how we should best understand and deploy AI. Our teams have benefited from the many rich conversations with our partners across the UK over the last few years, and this represents part of our response. It is also part of our ongoing commitment to transparency and integrity: an agency such as GCHQ should be able to explain to citizens how it is tackling the opportunities that surround emerging and novel technologies – while, of course, protecting its true operational secrets.

Defining AI should be straightforward – it is a subject that has been considered by academics and engineers since at least the 1950s.

What is Artificial Intelligence?

AI fact and fiction

Many academic articles and books have been published on the subject over the last decade alone, backed by a wide variety of university courses and studies.

In practice, there are numerous definitions, even among the expert community, which have continued to develop over time[5]. The Alan Turing Institute, the UK's national centre for data science and artificial intelligence, sums up AI succinctly as being about "machines which act intelligently". The Royal Society[6] recently defined AI as involving computers "learning directly from examples and experience in the form of data"; the Royal United Services Institute prefers to define AI as a technology that "enables machines to perform highly complex tasks effectively".[7]

For the purposes of this chapter, we will focus on AI in the context of tools which allow computers to tackle specific problems that require skills normally only available to humans: for example, analysing images, translating languages, manipulating objects, or making simple decisions.

Today, AI is already ubiquitous. As one commentator has observed:

"AI probably helped you get to work today. It's in our phones. In airports. In hospitals... artificial intelligence has become a ubiquitous part of modern society. Websites like Amazon and Facebook that tailor their content and recommendations are using interconnected networks of AI systems. Voice recognition. Search engines. Self-driving cars. All use AI."[8]

Despite that, the AI systems we can currently build are only able to tackle very limited, tightly-defined problems, and then only with human support. At least for the foreseeable future, AI engineers will certainly not be creating any intelligent computers that can truly replace humans. Computers with so-called "general intelligence", like the manipulative HAL from Stanley Kubrick's 1960s masterpiece 2001: A Space Odyssey, or the murderous Ava in Ex Machina, remain for now, firmly the stuff of science fiction.

Origins

We often think of AI as a purely modern phenomenon, but the mathematics underpinning it reach back several centuries. As early as 1801, for example, the German mathematician Carl Friedrich Gauss had invented a technique called "least squares regression" to predict where the asteroid Ceres would travel as it passed behind the Sun, removing the need for exhaustive work by fellow astronomers to find it again when it reemerged on the other side. His methodology remains a key part of AI software today.

Gauss, of course, did not have the benefit of modern computers and carried out his mathematical predictions by hand, but his British contemporaries Ada Lovelace and Charles Babbage were soon considering the construction of "intelligent machines" that could automate such processes. Babbage's colossal, mechanical "difference engine" – sadly unfinished – continues to inspire the imagination of many computer scientists, while Lovelace's contribution to data science and philosophy is commemorated through the Ada Lovelace Institute, and a GCHQ-sponsored Innovation Centre in Cheltenham.

A century later, the mathematician and Bletchley Park cryptanalyst Alan Turing and his colleagues were developing the foundations of digital computing and programming, which would ultimately make practical AI a reality. Turing was fascinated by the challenge of building computers that could demonstrate intelligence. In 1950,[9] he introduced the now well-known Turing Test: if a human being cannot distinguish a machine from a human in conversation, then that machine can be considered to be genuinely "thinking".

In the United States, academics such as Norbert Wiener drew on Turing's work to invent the new field of cybernetics, beginning the integration of digital computers and software into modern industry and everyday life. Wiener and his colleagues' work on communications and feedback mechanisms would prove essential to the development of AI.

But for many decades, AI theories ran well ahead of available technology: the mathematical ideas could often not be tested, let alone put into practice.

By the end of the 20th century, however, the digital revolution finally caught up with the theoreticians. Computer processors had become ever more powerful and efficient, backed by cheaper and larger storage capacity, and faster bulk data transmission over fibre-optic cables. This in turn enabled the rise of the new internet-based companies, feeding the demand for data and still greater computing power. AI was now a reality.

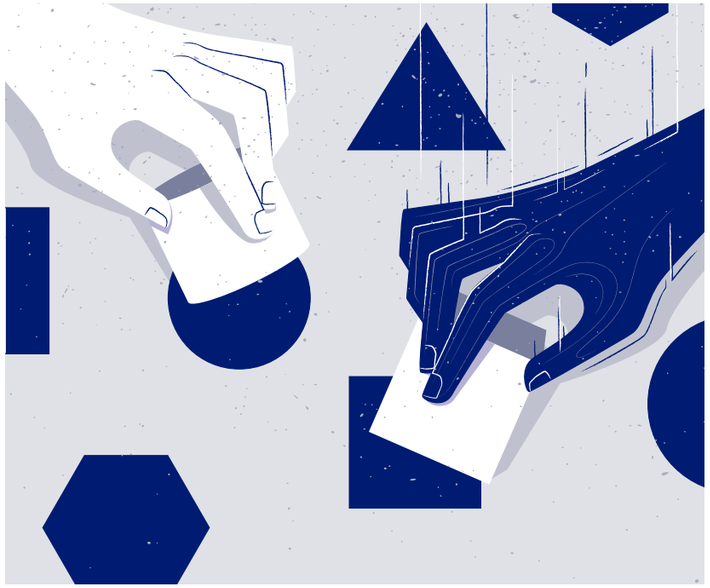
What is machine learning?

Designing AI systems requires a wide range of skills and academic disciplines, but perhaps the most important is an understanding of machine learning. This field of study has advanced hugely over the last few decades and underpins many of the major breakthroughs. But what is machine learning, and why is it so critical to AI?

All computers implement algorithms – mathematical instructions or rules applied to data – that help them calculate answers to a problem. When a piece of software works out the fastest driving route for you across the country, for example, it is using an algorithm. Algorithms can be very simple or extremely complex, but traditionally they have all had to be designed by humans, before being entered into a computer.

Machine learning enables us to automate that process. It allows computers to recognise patterns in data, understanding the relationship between the data they are given and the problem the designer is trying to solve, without it having to be explicitly programmed by a human. Rather than a human software designer trying to work out how a computer should calculate a good driving route, for example, machine learning enables the computer to discover the best method itself.

For this to happen, the machine learning software is given a set of pre-existing data, called "training data", and then cycles through a huge number of possible algorithms. The training data enables the software to evaluate the algorithms' effectiveness in solving the original problem, and guides it towards finding even better options.



Eventually, the machine learning software will have a preferred algorithm. Computer scientists call this result a "model" – the algorithm that offers the best solution for the problem. As the software developed and improved its model, we can think of it as having been "learning", albeit not in the same way a human would.

As a consequence, high-quality data is essential for machine learning. The algorithm that a machine learning package chooses as its final model is determined, in large part, by the training data it was set to learn on. It can be very hard to understand how a machine learning model works without understanding what training data it was originally given. In a way, we can think of the training data as having been "baked in" to the final machine learning algorithm used by the machine. As we will discuss in this paper, this has important implications for how we approach the challenge of AI ethics.

Different types of machine learning

Supervised Learning

There are various forms of machine learning, but one common approach is supervised learning, so-called because the training data that the software uses has already been analysed and assessed.

To take an example of this process in practice, a medical specialist might wish to take a set of scans of patients' retinas from health clinics and use machine learning software to quickly establish which of these patients are showing signs of serious disease, and are therefore most in need of urgent medical intervention. This will not replace the need for a human doctor – but it will get a specialist to the right patients more quickly.

To do this, the software would be given a collection of retina scans to learn from: its training data. This information will have been anonymised, analysed for signs of disease, and "tagged" with the correct diagnosis by a medical expert. As a result, the machine learning developer will know which of the images contain diseased retinas, and which were in fact healthy.

The machine learning software will analyse the images to predict which scan is showing signs of disease, testing its results against the real answers. It will do this repeatedly, steadily finding better and better algorithms. Eventually it will settle on the best model it has found, and pass that back to the developer.

As part of the validation process, the developer will then test the model on some fresh data – tagged retina scans the software has not seen before – to make certain the software is working correctly. When the developer and stakeholders are confident it is truly useful and safe to use, it can finally be deployed to help doctors prioritise real-world medical interventions.

Unsupervised Learning

In contrast, unsupervised learning involves using machine learning to make predictions based on data which has not been analysed and labelled by humans. This often involves the software looking for hidden patterns in large sets of data.

A cyber security analyst for example, may want to detect anomalous processes on a network or device – strange surges or peaks in digital activity that might indicate the presence of malware. Machine learning software can be used to analyse unlabelled training data from devices on the network, enabling the software to build a model to predict whether the given activity is normal or unusual. Future unusual activity would then be flagged up by the algorithm for expert human attention, enabling appropriate defensive measures to be taken.

The Deep Learning Revolution

One area of machine learning that has been particularly successful in recent years is deep learning, and its emergence has led to a rapid growth in AI products and services that had previously been considered too expensive or complex to build.

Deep learning involves creating artificial neural networks: complex networks of "neurons" which can process data and mimic the pathways found within our brains. A deep learning neural network is given large amounts of often quite simple data – such as the pixel information from an image file, or a sound wave from a microphone – and passes it through multiple layers of neurons. Eventually the data is processed sufficiently for the network to be able to analyse it correctly, for instance, recognising a diseased retina from an initial set of pixels in an image file.

Why do scientists call this "deep learning"? This phrase doesn't mean that neural networks are particularly profound, but instead just refers to the number of layers of neurons in the "deep network". Today's largest neural networks may have billions of neurons, arranged over hundreds of layers.

This complexity results in deep neural networks acting effectively as "black boxes". These systems do not need to be taught concepts or ideas to function, and can learn to analyse text and images without the need for bespoke coding. They can be thought of as "concept agnostic" and this is a significant strength as it makes them extremely versatile. But it also makes it hard, or impossible, to understand how those potentially billions of neurons made a particular prediction. To address this, the field of "explainable AI" is now developing techniques to help developers and users gain a greater understanding of how various AI methods process data and reach a conclusion.

Artificial neural networks were first developed in the 1980s, though without any commercial success. Their impact has been transformed over the past two decades, however, by two factors: greatly improved computational power – especially the use of specialised Graphics Processing Units – and the ability of internet and social media companies to harvest large, well-labelled training data sets from their customers' accounts. Deep learning techniques are notorious for needing lots of labelled data to learn from.

By 2010, researchers were demonstrating spectacular successes in using deep neural networks to recognise images and written characters. Deep learning was also applied to text and speech data, creating breakthroughs in text processing, speech recognition and natural language processing. Data scientists have proved that these networks can out-compete traditional methods in machine translation and they are now standard in the services we all use from our computers or smartphones.

Reinforcement Learning

Another machine learning approach is reinforcement learning, intended to enable AI systems to make decisions in changing and unpredictable environments.

While supervised learning learns from examples, reinforcement learning can be said to learn from experience. The machine learning software is programmed to make decisions which maximise a reward, typically the goal that the developer is trying to deliver. The software constantly considers the set of actions available to it, and tries them out in its operating environment, taking into account the level of reward which comes back in return.

Famously, in 2013, the British company DeepMind demonstrated a remarkable algorithm that was able to learn to play Atari video games, typically better than human players. The algorithm could only see the raw screen pixels, and the current game score, which it used as its reward signal. It was given no information about the concepts in the games – for example, that there were aliens, snakes, or tasks the player was trying achieve – and learned simply by operating the game controls and observing the impact on the screen pixels and its score, which it constantly tried to increase. The algorithm used a combination of reinforcement learning and deep learning, and DeepMind has since developed the concept much further, including defeating the world human champion at the Chinese game of Go.[10]

Applications of reinforcement learning go well beyond gaming, enabling advanced robotics, autonomous vehicles, logistics and infrastructure, cyber security and many other applications. Their only limitation is that the software needs to be able to experiment iteratively many, many times in order to learn – which in practice means that the problem they are tackling needs to be a digital, online challenge, or at least one that can be simulated accurately on a computer.

What AI can and cannot do

AI systems have proven themselves to be good at solving well-defined, narrow problems, where the necessary data and feedback are fully available to the system. When faced with this kind of task, AI systems are typically much faster and often more accurate than humans. AI systems are now able to perform tasks that would be so time-consuming for a human that they would otherwise be impossible to achieve.

But AI systems are not good at everything. Sometimes it can cost more to deploy an AI solution than a conventional software coding option, or simply getting a human to do a task in the first place. AI does not work well when tackling ambiguous, broad challenges, particularly if there is inadequate data on which it can train and learn. It faces problems in situations where the past does not predict the future well. And as we will discuss later in this paper, AI systems can be influenced by weaknesses in their data, incorporating biases, or proving possible to fool or mislead.

In addition, AI does not take into account the wider context, or many simple things which humans take for granted, such as psychology or emotion. For these reasons, most specialists argue that great caution should be exercised in applying AI for analysing individual behaviour. Deployed correctly, however, AI systems are a hugely powerful tool for any organisation or developer in the UK today.

"Machine learning....allows computers to recognise patterns in data, understanding the relationship between the data they are given and the problem the designer is trying to solve, without it having to be explicitly programmed by a human."

AI for national security

Cyber Threat

Almost half of UK businesses and a quarter of charities report having a security breach or cyber attack in the last 12 months, with one in five of these leading to significant loss of money or data.

Source: DCMS Cyber Security Breaches Survey 2020

GCHQ's role through its National Cyber Security Centre (NCSC) is to combat this threat and AI will be a vital tool in supporting this. Some examples of how it could do this include:

1

Helping us to identify malicious software by analysing patterns of activity on networks and devices at scale, characterising criminal or hostile behaviour, continuously updating our understanding, and correlating it with reported activity elsewhere.

2

In actively defending against cyber attacks, this knowledge will help us to spot malicious software faster by continuously looking for patterns of complex behaviours ("mining"), and by learning from these to update our known patterns (the "dictionary") to draw on to identify malicious software. It could do this by searching for anomalous events, such as particular website requests, blocks of unusual data, or suspicious emails.

3

Helping us to characterise and trace malicious software to its origin, in order to both support attribution and take down its sources.

When we are building new capabilities, this same knowledge will also allow us to accelerate discovery and remedy of vulnerabilities, as well as enable us to forestall new attacks against software by helping our industry to produce better products and services. We have been working with the Alan Turing Institute to develop a research roadmap for leveraging AI for cyber defence.[1]

The immediate cost of this can be measured in staff prevented from doing their work, lost revenue or damaged assets, including intellectual property theft.

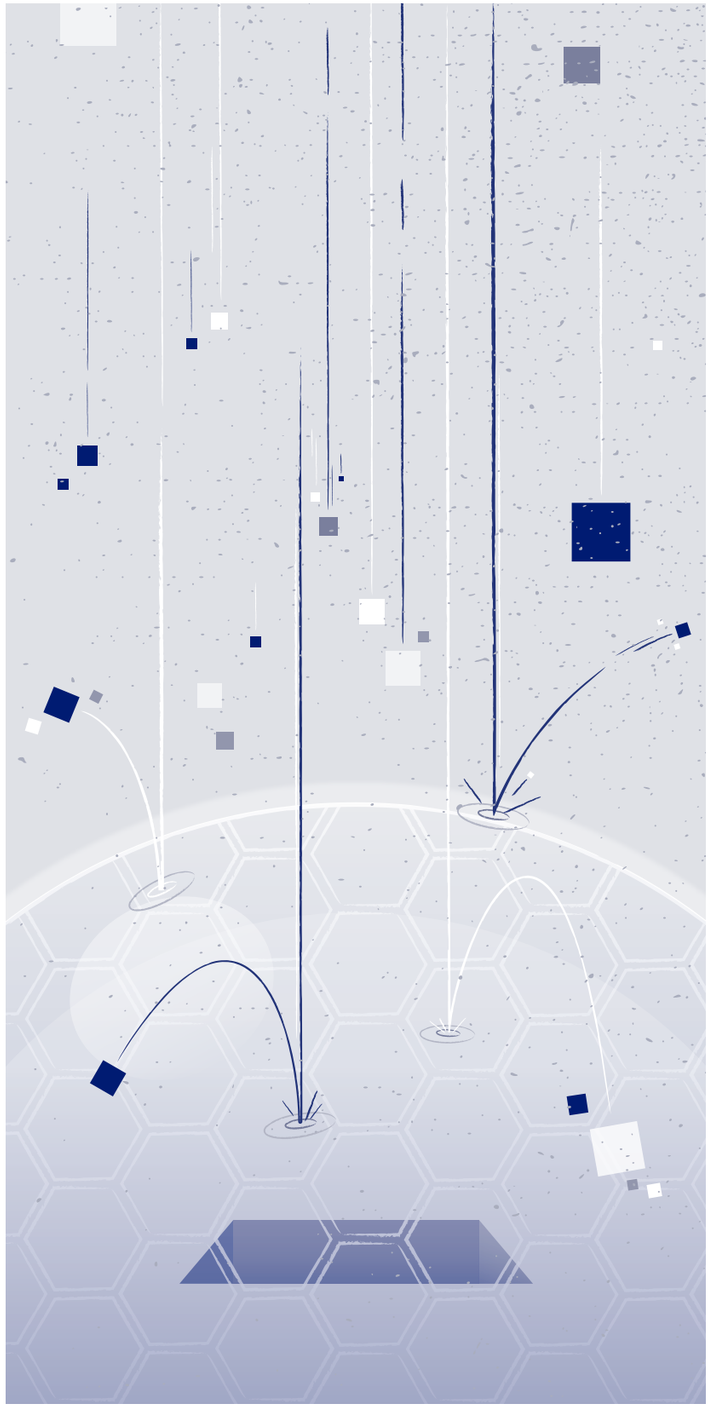
But there are also long-term costs such as loss of share value, investor and customer confidence, handling of complaints and payment of compensation, fines or legal costs.

Added to this can be the cost of remedy in repairing the vulnerability.

1. https://www.turing.ac.uk/research/publications/robust-artificial-intelligence-active-cyber-defence

What does GCHQ do?

GCHQ is a world-leading intelligence, cyber and security agency at the heart of Britain's national security community. We operate from hubs in three main locations: our headquarters in Cheltenham, our offices in central London, and our new team in the heart of Manchester.



Our extraordinary people use cutting-edge technology, technical ingenuity and wide-ranging partnerships to identify, analyse and disrupt threats.

It is GCHQ's job to stay one step ahead of those who would do our nation and its people harm. From managing the cyber threat posed by other nation states, to preventing terrorist attacks, keeping our children safe online and supporting our armed forces, the people of GCHQ operate on the frontline of global challenges. Part of GCHQ, the NCSC is helping to make the UK the safest place to live and do business online.

Our mission

We approach our work through the prism of our mission: keeping the UK safe. Our priorities are set by the UK's National Security Strategy[11] and the decisions of the National Security Council,[12] chaired by the Prime Minister, as well as through the direction of the Joint Intelligence Committee.[13] These priorities reflect our role under the Intelligence Service Act 1994: to protect the national security and the economic wellbeing of the United Kingdom, and to prevent and detect serious crime.

We carry out our mission through a number of unique contributions:

- Our intelligence meets the highest priority requirements through cutting-edge skills, covert global accesses and partnerships.
- The NCSC plays a leading role in making the UK the safest place to live and do business online.
- Our IT infrastructure and services are providing the next generation of national security capabilities.
- Our technology and policy expertise helps the UK to gain strategic technology advantage on all things data and digital.

We focus these unique contributions on five main mission areas:

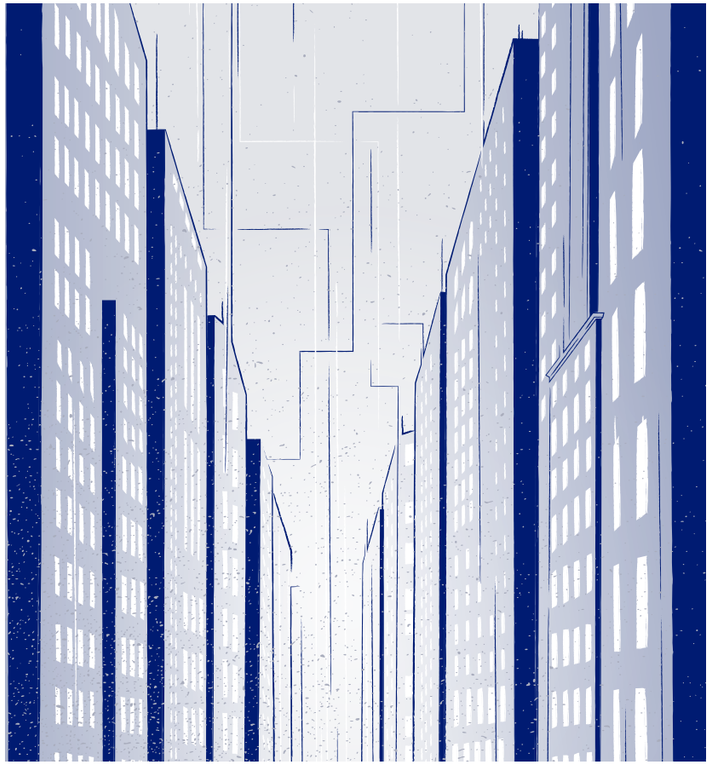
Counter terrorism. Working closely with MI5, the Police, Secret Intelligence Service (SIS) and the military, GCHQ plays a prominent role in keeping the UK safe from all forms of terrorism. We use our world-class capabilities to identify and disrupt terror plots against the UK and our allies. This work becomes more challenging as technology evolves, tactics change, and the internet grows, all of which requires the constant innovation of our staff.

We play a significant role in disrupting terrorist activity online. In recent years we have suppressed online recruitment propaganda such as that by Daesh[14] and hindered its ability to coordinate attacks on the battlefield and against our military personnel.

Cyber Security. The UK's cyber security mission is led by the NCSC. The NCSC helps to protect the UK's critical services from cyber attacks, manages major incidents and improves the underlying security of the UK internet through technological improvement and advice to citizens and organisations. It supports the most critical organisations in the UK, the wider public sector industry and small and medium sized enterprises. When incidents do occur, the NCSC provides effective incident response to minimise harm to the UK, helps with recovery and learns lessons for the future.

Strategic Advantage. We help to manage the threats to us and our allies from hostile states around the world. Our work also contributes to promoting the UK's prosperity and ensuring that the Rules Based International System is upheld. In recent years GCHQ has been involved in responding to the release of the online WannaCry attack by North Korean cyber actors, Russia's use of a nerve agent in Salisbury and Amesbury and a number of other attempts to disrupt our own and our allies' way of life.

Countering Serious Crime. GCHQ has a role in reducing the societal and financial harm which serious and organised crime causes within the UK. This involves us working closely with the National Crime Agency (NCA), HM Revenue and Customs (HMRC) and other government departments on a wide range of high priority topics such as cyber crime, Child Sexual Abuse (CSA) and organised immigration crime, including people trafficking. We work with our law enforcement partners to help them discover and disrupt criminal activity taking place in the real world or online. Our work also helps to bring these criminals to justice with many successful prosecutions from operations we have supported.



Support to Military Operations. Whenever and wherever British troops are deployed, GCHQ has worked hard to support their operations by ensuring they have intelligence to better understand the environment in which they are operating and to protect their personnel. Sometimes this is done remotely and at other times GCHQ people deploy along with the military units. In the recent conflicts in Iraq and Afghanistan, our team members have worked side by side with military colleagues to protect UK and coalition forces, with nearly 300 receiving campaign medals.

Our oversight

GCHQ's activities are governed by a range of legislative provisions which ensure our work is undertaken in accordance with the law. This legislation rigorously protects fundamental freedoms and essential human rights, while giving GCHQ the powers it needs to keep citizens safe and secure in the modern world. The Intelligence Services Act 1994 sets out our core functions as a global security and intelligence agency. The Investigatory Powers Act 2016 provides a framework that controls the use and oversight of investigatory powers by the security and intelligence agencies. GCHQ is also subject to the Human Rights Act 1998 which protects individuals' rights around the world, while the Data Protection Act 2018 lays out how we must approach the protection of personal data. Respecting fundamental rights is of central importance to GCHQ, especially those relating to privacy and the integrity of telecommunications systems. The Investigatory Powers Act 2016 provides that interception warrants will only be granted where:

- It must be **legally authorised** by a Secretary of State and approved by an independent Judicial Commissioner.
- It is **necessary** on certain limited grounds – meaning it seeks to achieve a legitimate aim in a democratic society, such as maintaining national security or countering serious crime.
- It is **proportionate** to the aim it seeks to achieve – which means there is no less intrusive way of achieving the objective, and that the need to achieve the outcome outweighs the impact of individuals' privacy or other rights.

We believe public trust is essential to any intelligence, cyber and security agency operating in a democratic country. We must not only operate within the law at all times, but should be able to demonstrate to Parliament and other independent bodies that we are doing so.

That transparency is provided through four key bodies:

- The Intelligence and Security Committee (ISC) provides oversight of our operations, policy, expenditure and administration to Parliament.
- The Investigatory Powers Commissioner's Office oversees the use of the powers we employ to conduct our operations. It is supported by the Technology Advisory Panel (TAP), which includes leading experts in new and emerging technologies.
- The Investigatory Powers Tribunal (IPT) is a judicial body that provides a route for redress for anyone who believes they have been the victim of unlawful action in our use of covert investigative techniques.
- The Information Commissioner's Office (ICO) oversees our compliance with relevant data protection legislation.

The laws and oversight that govern GCHQ's work are rigorous and world-class. The United Nations (UN) Special Rapporteur for Privacy, Joe Cannataci, observed following his review in 2018 that:

"The UK has historically been a force for good and a strong voice in support of the rule of law. UK legislative controls on Privacy and associated Codes of Practice…are amongst the most sophisticated and rigorous in the world."[15]

As well as the letter of the law itself, ethics matter too. As Director GCHQ stated in Singapore in 2019: "there are ethical rules and boundaries, and these should always be followed and upheld… our analysts are constantly reminded that it is not enough to be able to do something…it is not even enough for it to be legal to do something…it must also be right to do something."[16]

In 2014, GCHQ established an Ethics Counsellor – a Senior Civil Servant who gives specialist advice on ethics to staff at all levels, and keeps the GCHQ Board informed of any emerging concerns or issues, particularly around the use of new technologies.

As technology and society change and adapt, so must our legal and oversight frameworks. As Professor Sir David Omand, a former Director of GCHQ, recently noted: "public confidence is needed in the ethics of intelligence activity…[our] law needs to reflect a general consensus about what those ethical principles should be".[17] Frequent, well-informed debate is essential to this process. GCHQ welcomed two major reviews of the UK Investigatory Powers in 2015 and 2016 by Lord David Anderson, then the Independent Reviewer of Terrorism Legislation, and will be supporting the statutory review of the Investigatory Powers Act after its first five years of operation in 2021.

Our challenges and opportunities

In 2016, the UK Government published the "Operational Case for Bulk Powers", the first time that UK intelligence agencies had explained their work in such detail.[18] In that study, we observed how our society and economy was being radically transformed by developments in communications technology and computing.

The vast majority of these changes are extremely positive. The growth of the internet, mobile telecommunications and computing power, backed by strong encryption and effective cyber security, have brought huge benefits to UK businesses and citizens. They have enabled online commerce, increased international trade and created new business opportunities for the UK's growing information technology sector. The opportunities for the UK to prosper and grow are almost limitless.

But this transformation also fundamentally changed the operating environment for the security and intelligence agencies. Criminals, terrorists and hostile states were increasingly exploiting the weaknesses inherent to the internet, using the "dark web" and encryption to remain covert and anonymous. Older investigatory approaches were quickly losing their efficacy, it was proving harder to counter emerging threats, and the most vulnerable groups in our society were feeling the impact. In an ever more complex world, the UK needed a new kind of security.

Most of these challenges still apply today – alongside an ever growing number of opportunities to make the UK more prosperous and secure – in particular:

The rise of the data economy. The global internet continues to grow exponentially: the UN estimates that 4.1 billion people around the globe used the internet in 2019, approximately 700 million more than in 2016.[19] Almost every person in the world now lives within reach of a mobile phone signal.[20] Global data volumes are almost doubling every two years.[21] This growth is fuelling the fourth industrial revolution across our economy, transforming our lives for the better. The sheer volumes of data place huge pressure on security agencies and law enforcement, however, who must constantly adapt their analytic approaches to keep businesses and the public safe.

New technologies. Emerging technologies are reshaping the world around us. The internet of things (IoT), in which every object we use can interact with the global internet, or the roll-out of 5G telecommunications, enabling the development of "smart cities", represent a new generation of digital change. AI, the focus of this publication, is already embedded in our daily lives. We can expect the deployment of new computing techniques, synthetic biology and other emerging technologies over the next few years. Each new development helps our economy and society grow stronger, and provides opportunities to keep us secure, but also has the potential to be misused by those who seek to do us harm.

Skills and innovation. The UK is championing digital skills and bringing innovative services and products to market faster than ever before. Many of the old barriers between government departments focusing on economic growth and the agencies focused on national security have been broken down. For GCHQ, this has led to the creation of start-up accelerators and innovation centres around its three main hubs, including a new Innovation Co-operation Lab in Manchester, with the private sector learning from government and vice versa. Initiatives such as the NCSC's Industry 100 scheme have transformed our collaboration on cyber threats, while we have provided education and training to the next generation of technical leaders through the CyberFirst scheme.

As we look ahead to the next decade, GCHQ will need to constantly change and adapt in order to fulfil its mission to keep the UK safe. We expect AI to be at the heart of that transformation, which we will explore in the next chapter.

AI for national security

Online Safety

Child Sexual Abuse (CSA) is one of the most pernicious threats to our society.

GCHQ plays a vital role, with partners, in significantly reducing harm from CSA by reducing the volume and scope of online offending, targeting the highest harm and highest impact offenders and by creating a safer online environment. Some examples of how AI could assist us with tackling this threat include:

1

Providing tools and techniques to identify potential grooming behaviour within the text of messages and in chat rooms; highlighting the exchange of illegal images and tracking the disguised identities of offenders across multiple accounts; searching out and discovering hidden people and illegal services on the dark web. AI could also enable us to help law enforcement infiltrate rings of offenders and bring them to justice.

2

AI tools can also be trained to analyse seized and intercepted imagery, messages, other forms of internet content, and chains of contact, to support investigators in the identification of victims and discovery of accomplice offenders. AI running across both content and metadata could also protect our analysts from unnecessary exposure to traumatically disturbing material.

3

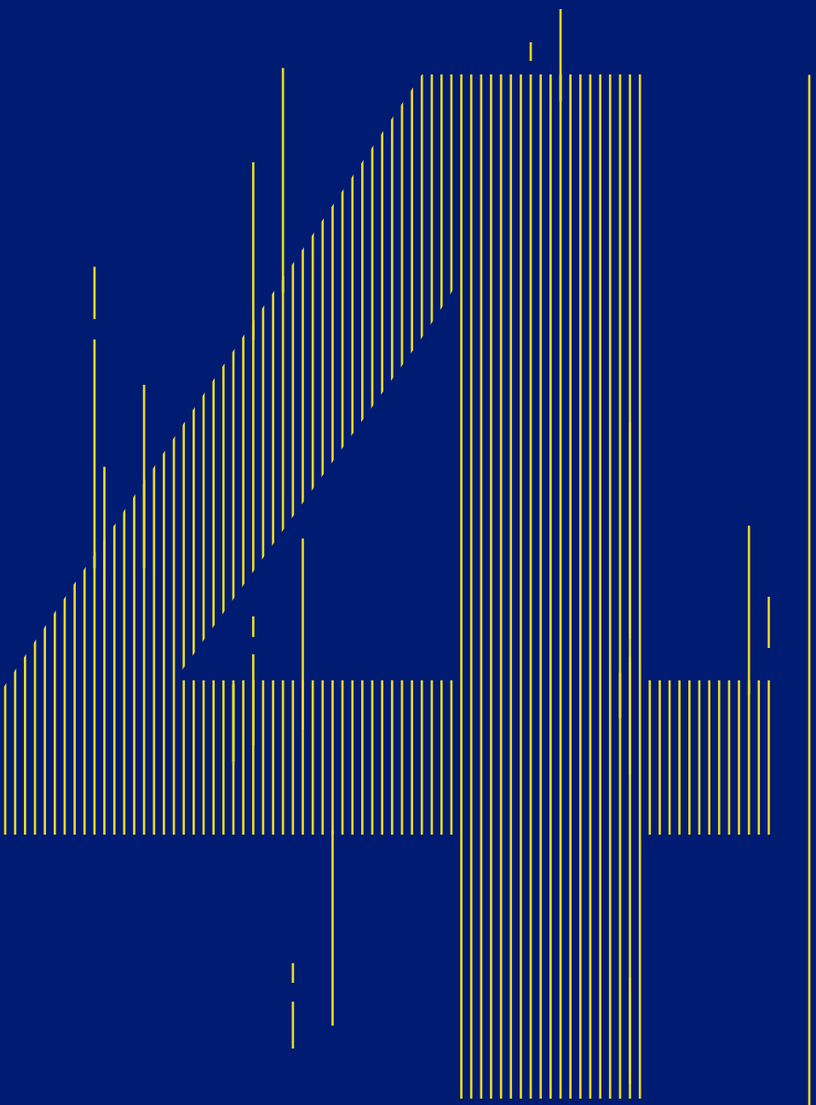
Enabling us to help other government departments, industry partners and charities understand and utilise AI technologies at scale to create a safer environment for children online.

Statistics from the National Crime Agency (NCA) show that in 2020, 2.88 million accounts were registered globally across the most harmful CSA dark web sites, with at least 5% believed to be registered in the UK[4].

Law Enforcement Agencies (LEAs) in the UK are each month arresting more than 500 individuals and safeguarding more than 700 children through their efforts to combat CSA[2].

In the UK alone, it is estimated that there are 300,000 people who present a sexual threat to children[3].

2 https://www.nationalcrimeagency.gov.uk/news/onlinesafetyathome
3 https://www.nationalcrimeagency.gov.uk/news/onlinesafetyathome
4 Home Office Factsheet on online Child Sexual Exploitation and Abuse 25.6.19



GCHQ has been in existence for just over a century and our history is inextricably linked with the development of data science and AI.

Artificial Intelligence in GCHQ

The invention of AI

The pioneering British cryptanalyst Alan Turing is widely credited with launching and inspiring much of the development and philosophy of AI with his 1950 paper "Computing Machinery and Intelligence". Turing had developed the principle of the modern computer in 1936 and he played a critical role in breaking ciphers at Bletchley Park during the Second World War. But this would not have been possible had he not been surrounded by a diverse team with varying expertise drawn from across UK industry and academia.

Post-war technology and computing

After 1945 many of Turing's team were dispersed, some to unsung and entirely different lives, but others to significant roles elsewhere. This included Donald Michie, who went on to become Director of the Department of Machine Intelligence and Perception at Edinburgh University and remained active in the research community into his eighties. The technological lessons of Bletchley Park began to be applied across the UK economy as a whole.

Michie's research was critical to the development of AI as it introduced the concept of increasing the probability of obtaining an accurate prediction – and as a result, increasing the effectiveness of an AI system – through a cyclical learning process. In 1960, at the time he developed these ideas, they were far beyond the ability of any computer to implement. Instead, Michie demonstrated the concept by implementing a program that could learn to play a perfect game of noughts and crosses using a collection of 304 matchboxes, each representing a unique board state. Each box was filled with coloured beads representing a different move from that state and the quantity of a colour indicated the level of certainty that playing the move would lead to a win.

Throughout the next decade advanced data techniques continued to be applied to complex industrial and administrative problems with practical applications. Soon, simple AI programs could be run on early digital computers. However, despite the Government and industry continuing to invest in computing, in 1965 there were still only 600 computers in the UK, with little expectation that the ambitions of Turing and Michie could become a reality. The "white heat" of the post-war scientific revolution began to cool.

The winter years

The period of the 1970s and 1980s is sometimes referred to as the "AI winter". Some predictions of the early commercial pioneers proved to have been ahead of their time, or simply overblown. Private sector investment faltered, many projects were cancelled, and academic funding often dried up. It was not until the development of the internet and the tools to enable it in the 1990s, that the thaw set in.

Nonetheless, data science continued to develop within GCHQ throughout the period. Our work on cryptanalysis attracted the very best mathematicians with backgrounds in statistics and probability. As in the Second World War, the challenges of working with large volumes of intercepted material continued to require specialists with strong data science skills, and sustained investment by successive UK Governments in computing power for the organisation built deep expertise in advanced programming. These skills, essential for the effective use of AI, were supported by our analytical culture and proficiency in telecommunications.

Our approach to delivering AI

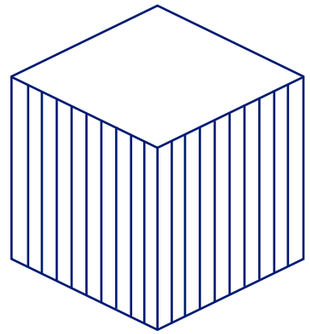
The last decade has seen a resurgence in AI technology, driven by a range of major technology companies and smaller, specialist developers. At GCHQ, we are combining these products and insights with our own specialist knowledge to develop new solutions for emerging security challenges.

This has required changes to our technological infrastructure. We are investing heavily in cloud technology and engineering, for example, to enable wider digital transformation and the use of AI. This is already yielding new operational capabilities and efficiencies and is driving organisational change, as our teams find themselves able to innovate more quickly than ever before.

Behind every technology however, are the technologists themselves. We have adapted our internal structures and processes to help recruit, train and develop some of the best data scientists in the UK. GCHQ has a growing community of data science and AI practitioners and researchers, including an industry-facing AI Lab dedicated to prototyping projects which address the applicability of AI to our mission. As this cadre of specialists grows, we are developing career structures and professional recognition for the discipline of data science – it is no longer a niche specialism, but a substantial part of our workforce.

Although we are proud of our role in the development of AI, most investment in the UK continues to come from the private sector rather than government and this is expected to continue. It is therefore unsurprising that GCHQ is now engaging more broadly with wider society and industry than at any other time in its history. We have much to learn from the exponential growth of AI in the outside world, and believe our specialists also have much to contribute.

That engagement spans both commercial partnerships and the development of oversight and ethics. We are fostering the UK's next generation of AI leaders, working in close partnership with local start-ups based around our hubs in London, Cheltenham and Manchester. These communities are developing diverse uses of AI, ranging from strategic forecasting to public health and safety applications. We are using our expertise and loaning our best people to inform the work of wider government. We supported the creation of the Alan Turing Institute in 2015 and are proud that it is named after one of our most prominent alumni – helping to join our history to the future direction of AI.



AI for national security

Trafficking

More than 350,000 individuals are estimated to be involved in serious organised crime (SOC) in the UK belonging to some 4,772 groups at huge cost to the UK economy.

Many SOC groups are involved in multiple types of trafficking, such as drugs, weapons and human trafficking and these in turn are enabled by other crimes; identity theft, bribery and the use of violence.

Their use of technology is becoming increasingly sophisticated, involving the use of encryption tools, the dark web and virtual assets, such as bitcoin and blockchain to conceal transactions.

Source: National Strategic Assessment of Serious and Organised Crime 2020

GCHQ's role in the prevention and detection of serious crime is to support the National Crime Agency and Law Enforcement Agencies in combatting this threat. Some examples of how AI could assist us with this include:

1

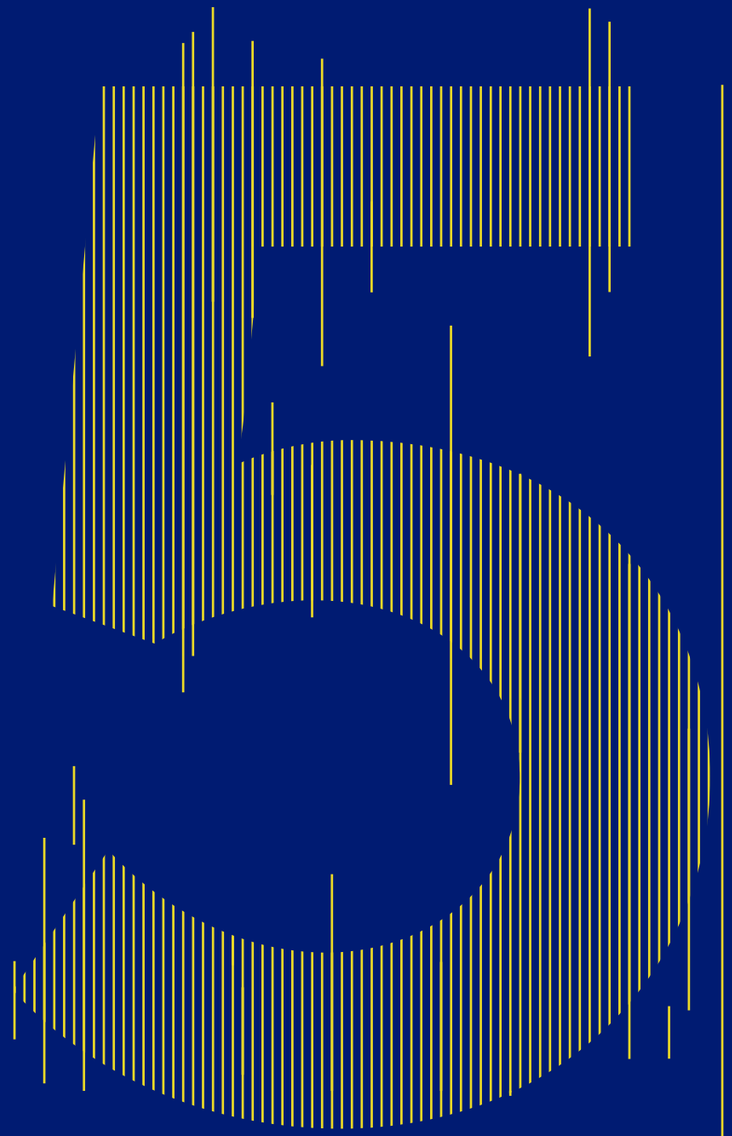
Helping us to map the complex international networks that enable trafficking – identifying individuals, accounts and transactions to reveal criminal groups and their associations.

2

Assisting us to "follow the money" – analysing large scale, complex chains of financial transactions as payments are made and received online, possibly even revealing state sponsors or terrorist associations.

3

AI tools could provide geographical information on illicit activity – enabling the analysis of multiple sources of imagery, messaging, sensor data and other information to track and predict the delivery of illegal cargos.



The ethical development and deployment of AI has become a strategic challenge for governments around the world.

GCHQ, AI and trust

International and national context

Last year, the Secretary General of the United Nations observed that:

"Artificial intelligence brings the promise of improved access to healthcare, accelerated economic development and other gains. But we can also see dangers: a world with diminished privacy, less human agency and accountability, and where income inequality widens and access to work narrows for millions…If we are to harness the benefits of artificial intelligence and address the risks, we must all work together – governments, industry, academia and civil society – to develop the frameworks and systems that enable responsible innovation."

The UK has been at the forefront of the international debate around ethical and responsible approaches to AI, reflecting our leading role in AI research and development. As the Prime Minister explained to world leaders in New York in 2019, our challenge is to ensure that AI is "designed from the outset for freedom, openness and pluralism, with the right safeguards in place to protect our peoples."

At GCHQ, we are committed to helping to deliver against this commitment. The research of the UK Centre for Data Ethics and Innovation (CDEI) informs our approach and we are a Strategic Partner to the Alan Turing Institute, providing expert support to its world-class research programme. The NCSC leads UK efforts to improve the cyber security protecting AI systems, both by deploying existing best practices and by developing new methods to counter specific vulnerabilities unique to AI-based systems. Internationally we are working with the European Telecommunications Standards Institute (ETSI), which is seeking to develop stronger global standards.[22]

We believe that rigorous debate between groups with diverse perspectives and experiences will be critical to resolving the challenges surrounding the ethical use of AI. In recent years, we have met with academic researchers specialising in AI, civil society bodies championing privacy rights and data privacy, and thought leaders from across the private sector. These discussions have informed GCHQ's publications on AI and ethics in various specialist journals, and our engagement at conferences in the UK and Europe. Finally, we funded a major independent review into the role of AI in national security, carried out by the UK's Royal United Services Institute (RUSI), the results of which were published in April 2020.

What is AI ethics?

The Alan Turing Institute explains that the field of AI ethics emerged from the need to address the individual and societal harms AI systems might cause. These issues rarely arise as a result of a deliberate choice, as most AI developers do not want the applications they build to be biased or discriminatory, or to invade users' privacy. Instead, the problems found around AI systems are most commonly a consequence of:

- **Mis-deployment –** software being used for purposes other than for which it were originally designed.
- **Questionable technical design –** technical risks related to bias and safety not being fully resolved by designers.
- **Unintended negative consequences –** the potential for systems to cause harm to individuals or communities not being foreseen or tackled during development.

Our own experience in protecting the high quality, secure data needed for AI systems, highlights that problems may also result from:

- **Poor cyber security –** software being attacked by adversaries intending to affect the behaviour of a system, or to gain insights about its users.

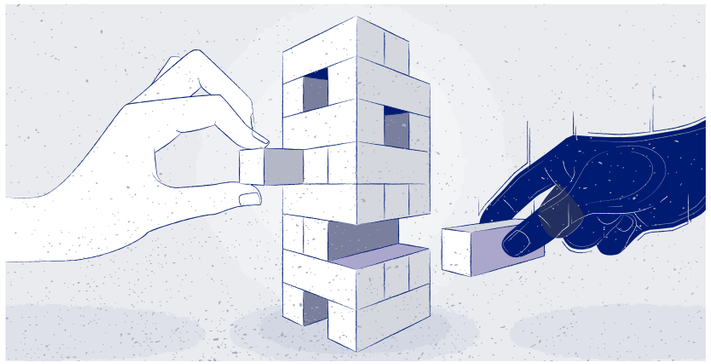
The field of AI ethics seeks to mitigate these risks by providing developer teams with the values, principles, and techniques needed to produce and use ethical, fair, secure and safe AI applications. The challenge for GCHQ is to manage these risks, which apply to our own work as much as any other organisation, while operating in the unique – often highly classified – environment of national security. Our adversaries work in secret, and often, so must our teams – but we must still be as open as possible about our approach to AI ethics, maintaining trust, and learning from others' experiences and insights.

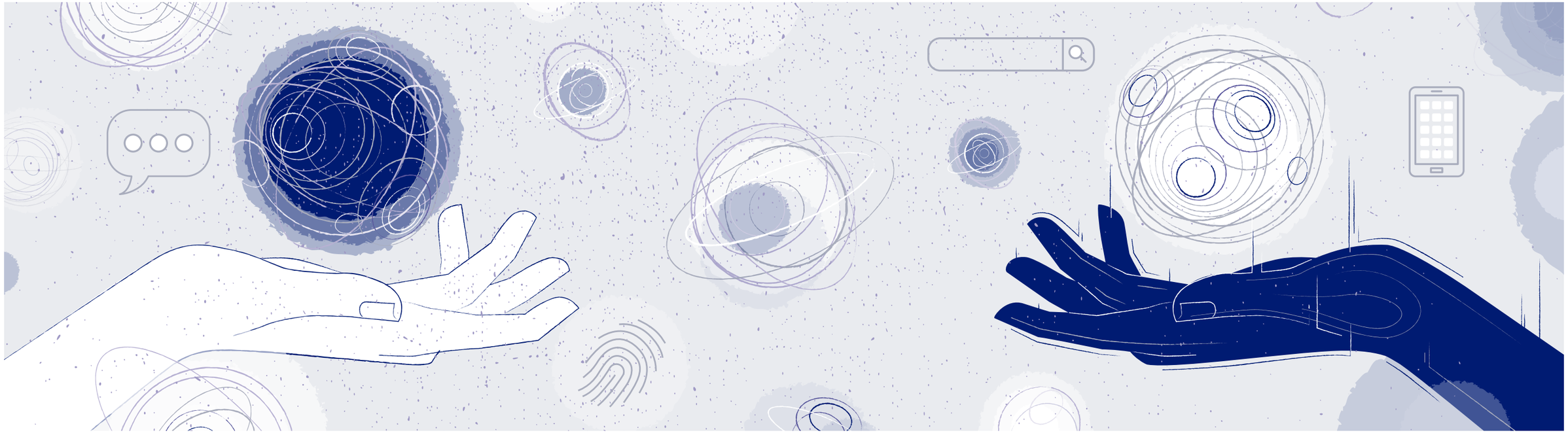
AI, ethics and National Security

GCHQ's approach to AI ethics is informed by the framework generated by the Alan Turing Institute, published as "Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector" in 2019. It aims to help organisations create ethical AI capabilities, which are fair and non-discriminatory, worthy of public trust, and justifiable to stakeholders.

This framework captures best practice for both governmental and non-governmental bodies. It urges the maintenance of a deeply ingrained culture of responsibility, and execution of ethically sound practices at every point in the AI innovation and implementation lifecycle.

GCHQ's design and implementation of AI systems also takes into account the findings from the recent RUSI research project into national security and AI. Recognising that the nature of these ethical challenges sometimes needs to be considered specifically in a national security context, this research is providing an independent evidence base to further inform our thinking.





The major ethical challenges

FAIRNESS

Fairness – that sense of equity and reasonableness we all know when we experience it, but probably would find difficult to perfectly define – is an essential attribute of any good system. Fairness, or more commonly its antithesis, bias, has been extensively studied by AI ethics researchers. The range of challenges can include:

- **Data fairness.** Many organisations have found that they have unknowingly trained supervised AI software on skewed data sets, only to find that the system then either works better in supporting particular groups – typically, white men – or actively discriminates against others, most often women or those from an ethnic minority background. Some of the more prominent examples have included speech and facial recognition software that fails to work appropriately for large parts of society.
- **Design fairness.** Any AI designer has to make choices in creating a new system, including deciding what sorts of data to include for the software to analyse – even if the system uses unsupervised learning. The decisions we make reflect our own assumptions and biases; we may assume that ethnic background, age, or social class are important factors, for example, and prioritise feeding those into a model. Those decisions in turn can make an AI system unfair or unethical.
- **Outcome fairness** Moving beyond our technical decisions, an AI system can also be biased if it fails to treat individuals equally, or has an impact on particular groups that is unreasonable. If we give an AI system an unacceptable goal, perhaps driving unsupported assumptions of the ethnicity or gender of a target, it may well learn how to achieve it, but reinforce social discrimination in the process. This may be a result of technical design, or how a system is actually used in practice, but the impact on an individual will be just as significant.

These risks are all applicable to the national security context, particularly where our analysts are seeking to derive insights about individuals. Best practice in managing these risks involves proactively addressing fairness from the very start of an AI project; documenting the characteristics of what "fairness" might look like; and monitoring it throughout the lifecyle of the project.

At GCHQ, we will be able to demonstrate that we have implemented measures to minimise bias, and that we have scrutinised our AI models' outputs to mitigate the potential for discriminatory outcomes. We have extensive expertise in applying complex analytics and mathematical models to large datasets, and will utilise both this and wider best practice to develop this approach.

We believe that true fairness goes further than this. We concur with Professor Wendy Hall and Jérôme Pesenti, when they argue that if software developers are not representative of the nation as a whole, their capability and credibility to tackle AI bias will be limited, and with the Committee on Standards in Public Life which has stressed that: "a workforce composed of a single demographic is less likely to check for and notice discrimination than diverse teams."[23]

Artificial Intelligence at GCHQ

EMPOWERMENT

Effective AI has the potential to empower human decision-makers, providing them with information and insights they might otherwise have lacked, expanding and improving their available options. Conversely, poorly designed or implemented AI can disempower, removing human agency from a process or masking the role of the software in decision-making. Some researchers describe this challenge as being one of keeping a "human in the decision loop", reflecting our natural reluctance to cede decision-making entirely to machines. For GCHQ, empowering our people to manage and benefit from AI starts with the basics: good education in both how AI works and the principles of AI ethics. All of our people will have the right opportunities to learn, and we will provide additional support for key groups, including our senior leadership.

Keeping humans in the decision loop is nonetheless particularly challenging for organisations and teams that operate primarily in the online world. As we have previously discussed, many of the operational areas where we most need AI support have to operate at machine speed if they are to be effective in protecting the UK. Cyber defence systems designed to use AI predictions to prevent potential cyber attack, for example, often need to respond near instantaneously when they detect threats. Few analysts would want their firewall system to delay putting any protective measures in place until a human could be found to confirm their decision.

GCHQ's specialists share the same concerns voiced by many external experts around using AI to make predictions about individuals, their behaviour and motivations. AI software can help triage and prioritise across our data sources. It can suggest previously unseen patterns and learn to identify valuable behavioural indicators. But it is not yet sophisticated enough to be trusted to make independent decisions based on those outputs.

In these cases, we expect our approach to AI to resemble the augmented intelligence model being advocated by a range of research partners, including RUSI. This involves tasking AI software to collate information from relevant sources and flag significant conclusions for review by a human analyst, but does not automate any action as a result – it supports and empowers the human decision-making process rather than determining it.[24]

The major ethical challenges

TRANSPARENCY AND ACCOUNTABILITY

Much has been written regarding the way in which AI software makes predictions, and whether this can be adequately explained and understood. With machine learning, the computer itself uses training data to develop and improve its predictions. Using some advanced AI methodologies, in particular deep learning, it will not always be possible for a human to fully assess the factors that the software took into account to form its conclusions.

As previously discussed, this is sometimes termed the "black box" problem, and has implications for accountability in the overall decision-making process. This is of particular relevance for an organisation such as GCHQ, where decisions could impact the privacy of individuals or the security of the nation – we have to be able to explain how we make our decisions. The problem is also more significant because such AI techniques are often amongst the most powerful and effective available to our developers.

To ensure that we can account for our decisions, it is important for our testing procedures to demonstrate the validity and reliability of the AI method employed. Where appropriate, we will use the developing techniques emerging from the field of "explainable AI" to improve our assurance. It will be essential that we design the systems in such a way that non-technically skilled users can interpret key technical information such as the margins of error and levels of uncertainty. Where the systems are drawing conclusions about individuals, this analysis should form part of a wider evidence base from which our analysts can make an overall judgement.

Performance of an AI system can also drift and decline over time. Changes in the external environment may mean that the data on which it was originally trained is no longer optimal for the task it now faces – an AI system that was effective in spotting computer malware several years ago is unlikely to be as effective today, for example, unless updated and adapted. We will therefore ensure that a lifecycle plan for each AI system is recorded to assure its long-term health, including its security.

PRIVACY

The right to a private life is a fundamental human right. It is a key part of the Council of Europe Convention for the Protection of Human Rights and Fundamental Freedoms of 1950 and the Human Rights Act 1998, which underpins all of the legislation that governs GCHQ's work. We recognise that many of our operations will impact on privacy rights to some degree, regardless of where in the world they may be taking place. That is why our legal framework mainstreams human rights considerations into all investigatory powers, regardless of whether the legal instruments described above are technically applicable as a matter of law. Moreover, the legal framework that Parliament bolstered and updated through the Investigatory Powers Act 2016 aimed to be "technology neutral" – precise in its principles and safeguards but adaptable when it came to the technological means by which GCHQ and others could accomplish their aims.

There are a range of views regarding AI software and the impact it may have on privacy rights. Some researchers argue that the greater use of analytic systems, supported by machine learning, could have a "chilling effect" on individual privacy or the rights of sensitive groups. Some fear that multiple AI systems may interact with each other to create a greater "cumulative intrusion risk" than might be the case for each system in isolation. Others conversely argue that AI approaches will actually reduce the level of intrusion into privacy, because the software will minimise the volumes of personal data that will need to be reviewed by a human analyst. Some focus on the benefits of AI for protecting our personal data from cyber criminals or other malevolent actors.

To ensure the impact on privacy is properly considered in every circumstance, GCHQ undertakes an assessment of the necessity and proportionality of any intrusion into privacy both when considering the use of operational data to train and test AI software, and when applying the software to the analysis of operational datasets. The assessment of necessity and proportionality is already integral to the activity our people undertake:

Necessity. This requires the analyst to ensure that the activity is necessary to meet a legally valid intelligence requirement which falls under one of our statutory purposes as set out in the Intelligence Services Act 1994, and there is no reasonable prospect of obtaining the wanted information through other means.

Proportionality. Ensuring proportionality requires the maintainance of a justifiable case-by-case balance between the intrusiveness of what is planned and the value that would be derived, whilst representing the minimum interference necessary to achieve it.

To ensure we meet our legal obligations and compliance standards, including on rigorous necessity and proportionality assessments, all of our people must adhere to detailed internal compliance rules. The statements we provide in relation to operational activities governed by the Investigatory Powers Act are subject to regular audit, inspection and oversight by IPCO, and all of our people undertake comprehensive training in relation to their responsibilities and how to meet them.

Our approach - next steps

We are developing a comprehensive governance system to manage AI and data ethics, which will continue to ensure the necessary safeguards are in place to safely implement AI in GCHQ. It will consist of the following:

An AI Ethical Code of Practice which draws on best practice around data ethics, and builds systematically on the practical experience we are acquiring as our specialists continue to develop capability. It comprises an internal policy, setting out the standards our software developers are expected to meet, and supporting guidance explaining the techniques and processes which can be employed to achieve this. We have been managing complex analytic systems for over a century and intend to use the full weight of this accumulated knowledge and experience to manage our approach to AI ethics.

World-class AI training and education. All of our people have a personal responsibility to understand and comply with the legal and ethical obligations placed on GCHQ. To support them, we will deliver training and education in the issues and challenges that AI raises for all levels of the organisation. We are strengthening the way we cultivate and grow our data scientist professionals, and are investing in the specialist training of those engaged in the development, use and security of AI systems. The skills and awareness of our people will be reflected in the quality of the systems we design, build and operate.

The right "mix of minds". We will monitor the way we grow our teams and are committed to reflecting the nation we serve. This means drawing on the full spectrum of talent from across the UK to drive innovation and ingenuity underpinned by our values. We foster a culture of challenge, proactively seeking alternative perspectives and ideas; pushing ourselves to continually question assumptions. We know that we still have much to do to achieve the right mix of minds, but we recognise that the UK's diverse talent pool is one of our key assets, providing resilience and the ability to respond to new challenges. Although we must retain much of our work behind a protective shield of security, we remain an outward looking organisation, ready to learn from best practice wherever we might find it.

Reinforced AI governance. We are reviewing our internal governance processes to ensure they apply for the full lifecycle of an AI system. This includes an escalation mechanism for the review of novel or more challenging AI applications.

In all we do we will be guided by the GCHQ core values:

INTEGRITY
INGENUITY
IMPACT
TEAMWORK

Always use our AI in a legal, proportionate and ethical way;

Explain publicly how and why we use AI, while protecting our genuine secrets;

INTEGRITY
INGENUITY
IMPACT
TEAMWORK

Draw on the very best of our people's AI experience, technologies and tradecraft;

Resolve AI challenges and unlock national opportunities that others thought impossible;

Artificial Intelligence at GCHQ

INTEGRITY
INGENUITY
IMPACT
TEAMWORK

Use our AI to protect the UK – making a real difference to the nation's interests;

Take the difficult decisions – both around when we decide to deploy AI, and when we choose not to;

INTEGRITY
INGENUITY
IMPACT
TEAMWORK

Seek the right mix of minds and foster an inclusive approach to our AI development;

Build fairness into everything we do, from AI design through to implementation.

AI for national security

Foreign State Disinformation

Hostile actors can use AI to mount disinformation attacks by automating the production of false content to undermine public debate, including the production of "deepfake" video and audio material designed to mislead.

It can also be used to inject fake personas into debate through the use of AI chatbots.

AI has also been known to be deployed to manipulate information availability through interference with content curation algorithms.

AI analysis can be used to target individual user profiles with tailored information to enable personalised political targeting.

A growing number of states are using AI-enabled tools and techniques to pursue political ends by spreading disinformation to shape public perceptions and undermine trust. But AI could also assist the UK in tackling this threat:

1

In defence against these techniques, AI enabled tools could be deployed for machine-assisted fact checking through validation against trusted sources and to detect deepfake media content.

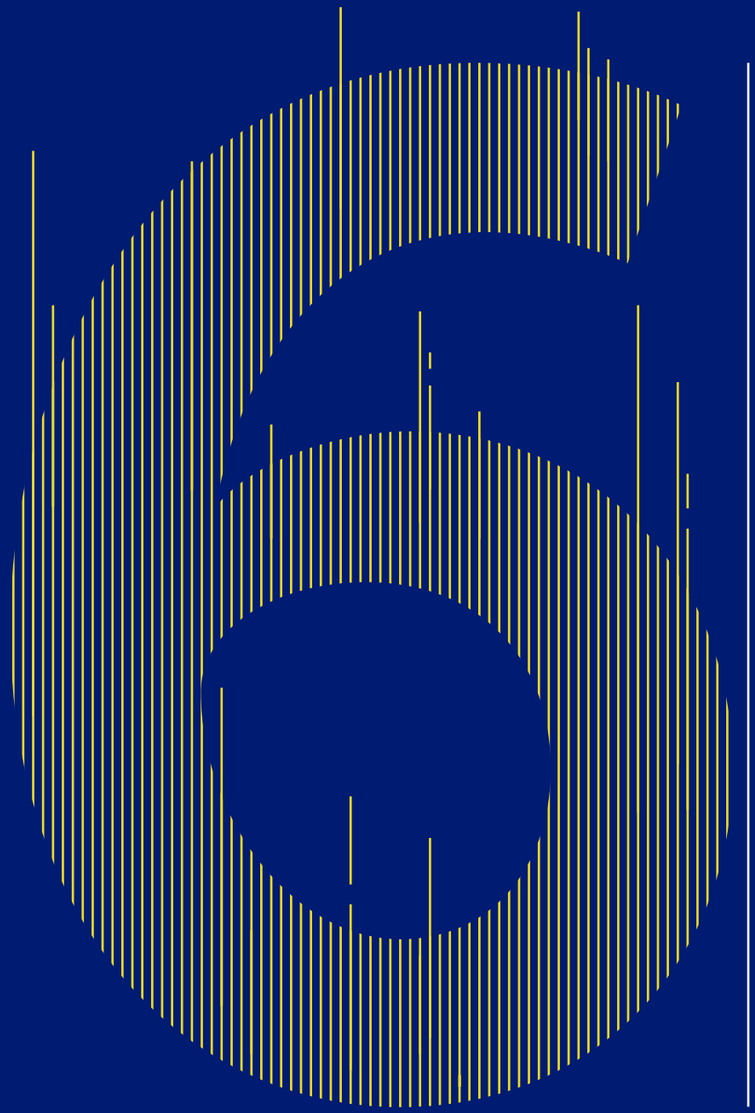
2

Providing us with the techniques and ability to detect and block botnets[5] with machine-generated social media accounts.

3

Helping us identify the "troll farms"[6] and sources generating the misinformation in the first place, so that online operations could be mounted to counteract these malicious accounts.

5 a network of private computers infected with malicious software and controlled as a group without the owners' knowledge, e.g. to send spam.
6 an organisation employing people to make deliberately offensive or provocative online posts in order to cause conflict or manipulate public opinion.



GCHQ has come a long way since the days of Bletchley Park, Alan Turing and the first Colossus computer. We stand on the shoulders of those giants as we build security and intelligence capabilities of which our predecessors could only have dreamt.

Conclusions and our future journey

As we have discussed, the exponential growth of global computing capacity and data volumes, combined with the brilliance of countless scientists and developers, has enabled the development of what we now think of as AI – powerful software algorithms that are transforming the world in which we all live.

This paper has outlined our values-led ambition and the potential for AI to help organisations like GCHQ deliver their mission. But in democracies such as our own, we cannot do that without embracing AI ethics – ensuring teams have the values, principles, and techniques needed to produce and use ethical and safe AI applications.

Britain's diplomats, academics and campaigners have led much of the international debate around technology and values. GCHQ is using the groundbreaking work by the Alan Turing Institute to shape our own AI Ethical Code of Practice – building AI systems that embrace fairness, empowerment, transparency and accountability. Enabled by the UK's world-class legislation governing the use of investigative powers and the protection of citizens' data, we believe that we can safely use AI in the heart of the intelligence security community, and protect the nation's wider AI systems from attack and disruption. We hope that some of the operational examples in this paper have explained why this is so important to the nation.

The veterans of Bletchley Park worked under conditions of the utmost secrecy; as Winston Churchill remarked, they were the "geese that never cackled". Today, we are committed to being much more open in engaging with wider British society about the technologies and choices that we make on behalf of the nation. We hope this paper has provided you with some insight into our approach and we look forward to continuing the conversation.

Endnotes / Citations

1 *Hall, W. and Pesenti, J. (2017). Growing the artificial intelligence industry in the uk.* [online] Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/652097/Growing_the_artificial_intelligence_industry_in_the_UK.pdf

2 Nations, U. (2019). *Towards an Ethics of Artificial Intelligence | United Nations*. [online] Available at: https://www.un.org/en/chronicle/article/towards-ethics-artificial-intelligence

3 *Hall, W. and Pesenti, J. (2017). Growing the artificial intelligence industry in the UK.* [online] Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/652097/Growing_the_artificial_intelligence_industry_in_the_UK.pdf (Growth rate measured through Gross Value Added (GVA).

4 www.gchq.gov.uk. (2018). *Director GCHQ writes about the importance of securing the next generation of technology.* [online] Available at: https://www.gchq.gov.uk/news/jeremy-fleming-securing-next-generation-technology

5 Marr, B. (2014). *The key definitions of Artificial Intelligence (AI) that explain its importance.* Forbes.

6 *Machine learning: the power and promise of computers that learn by example 1*. (2017). [online] Available at: https://royalsociety.org/~/media/policy/projects/machine-learning/publications/machine-learning-report.pdf?la=en-GB&hash=B4BA640A1B3EFB81CE4F79D70B6BC234

7 RUSI. (2020). *Artificial Intelligence and UK National Security: Policy Considerations.* [online] Available at: https://rusi.org/publication/occasional-papers/artificial-intelligence-and-uk-national-security-policy-considerations

8 The Alan Turing Institute. (2018). *What does AI mean for the Turing?* [online] Available at: https://www.turing.ac.uk/research/research-programmes/artificial-intelligence-ai/programme-articles/what-does-ai-mean-turing

9 Alan Mathison Turing (1950). *Computing machinery and intelligence*. Aberdeen Univ. Press.

10 Go is a board game like chess, played with black and white stones. It has much simpler rules than chess but many more possible moves, and is a much greater computational challenge than chess.

11 Prime (2015). *National Security Strategy and Strategic Defence and Security Review 2015.* [online] Available at: https://www.gov.uk/government/publications/national-security-strategy-and-strategic-defence-and-security-review-2015

12 GOV.UK. (2016). *National Security Council*. [online] Available at: https://www.gov.uk/government/groups/national-security-council

13 GOV.UK. (n.d.). *Joint Intelligence Committee*. [online] Available at: https://www.gov.uk/government/groups/joint-intelligence-committee

14 A name sometimes used to denote the Islamic State of Iraq and the Levant (ISIL)

15 End of mission statement of the special rapporteur on the right to privacy at the conclusion of his missions to the United Kingdom of Great Britain and Northern Ireland

16 www.gchq.gov.uk. (2019). *Director's speech on Cyber Power - as delivered.* [online] Available at: https://www.gchq.gov.uk/speech/jeremy-fleming-fullerton-speech-singapore-2019

17 Omand, D. and Phythian, M. (2018). *Principled spying : the ethics of secret intelligence*. Oxford: Oxford University Press.

18 *Operational Case for Bulk Powers.* (2018). [online] Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/504187/Operational_Case_for_Bulk_Powers.pdf

19 Itu.int. (2015). *Statistics.* [online] Available at: https://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx

20 Itu.int. (2019). *Press Release*. [online] Available at: https://www.itu.int/en/mediacentre/Pages/2019-PR19.aspx

21 Un.org. (2018). *Big Data for Sustainable Development.* [online] Available at: https://www.un.org/en/sections/issues-depth/big-data-sustainable-development/index.html

22 Dahmen-Lhuissier, S. (2020). *ETSI - Best Security Standards | ETSI Security Standards.* [online] Available at: https://www.etsi.org/technologies/securing-artificial-intelligence

23 GOV.UK. (2020). *Artificial Intelligence and Public Standards:* report. [online] Available at: https://www.gov.uk/government/publications/artificial-intelligence-and-public-standards-report.

24 RUSI. (2020). *Artificial Intelligence and UK National Security: Policy Considerations.* [online] Available at: https://rusi.org/publication/occasional-papers/artificial-intelligence-and-uk-national-security-policy-considerations

Artificial Intelligence at GCHQ

Pioneering a New National Security

The Ethics of Artificial Intelligence