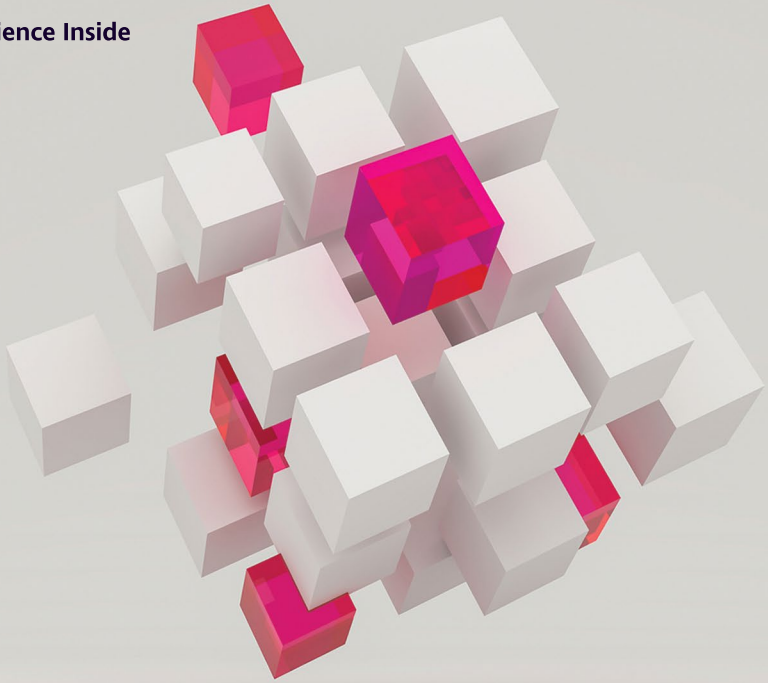
[dstl]

The Science Inside



Defence Science and Technology Laboratory

Building Blocks for Artificial Intelligence and Autonomy

A Dstl Biscuit Book

Ministry of Defence

Defence Science and Technology Laboratory

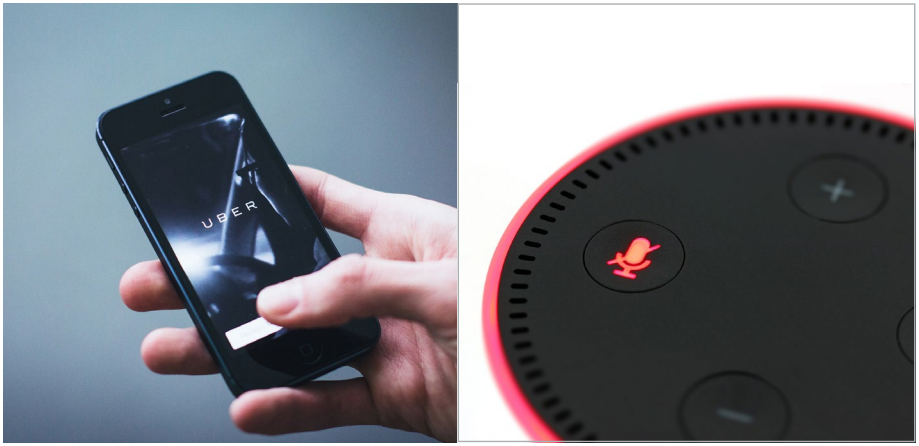
Building Blocks for Artificial Intelligence and Autonomy

A Dstl Biscuit Book

DSTL/PUB126301 v1.0

Contents

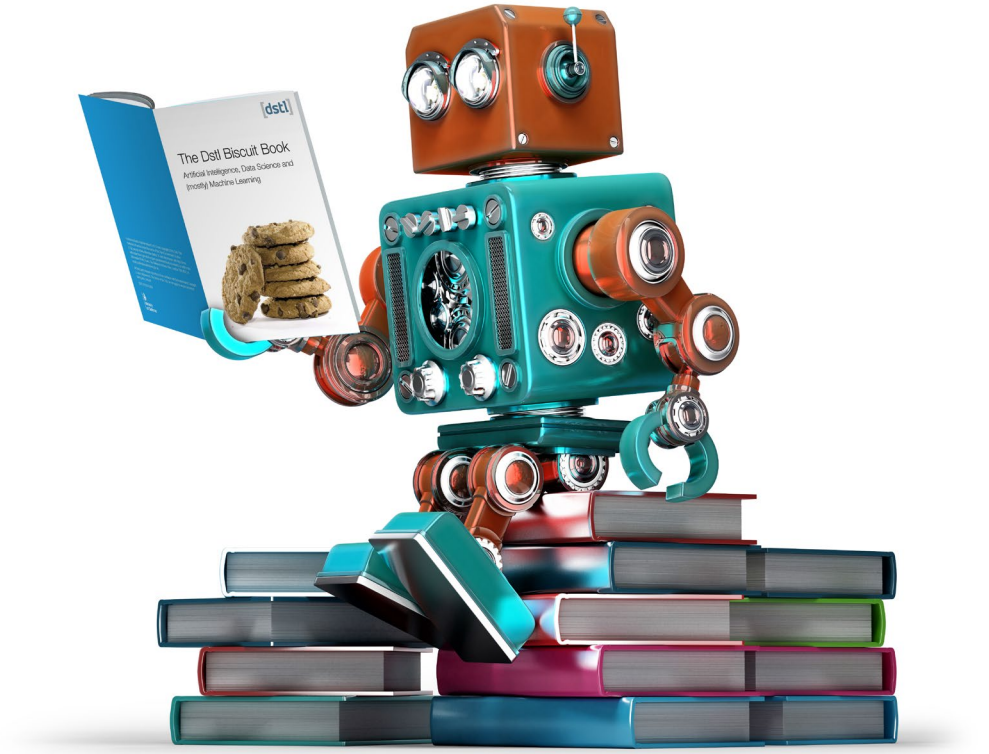
Foreword



Artificial Intelligence (AI) and Autonomy are experiencing a surge in interest across almost all sectors of the global economy, offering breakthroughs in many areas previously beyond the capabilities of traditional approaches to computing. These terms, so long confined to academia, have entered the public conscience, through consumer applications like Amazon's Alexa, Google's Translate, and Uber's self-driving cars, and business applications, like Ocado's retail technology.

Despite these many popular examples of AI and Autonomy, actually realising the benefits is deceptively hard. We often see compelling demonstrations of the technology and assume it can quickly bring value within our own areas. However, this is rarely the case. Too often, we focus on the technology in isolation, ignoring the other critical elements needed to realise the benefit and deliver a genuine new capability.

To address this challenge, we have developed the Building Blocks for AI and Autonomy. We aim to provide a framework for thinking about AI and Autonomy from a systems perspective, and provide some practical questions you should consider when undertaking AI and Autonomy projects. So boil the kettle, grab a biscuit and get reading ...



This publication is intended to stimulate thought and provide advice that is based on the cumulative experience of the S&T teams delivering MOD funded AI research projects; it does not represent departmental or government policy.

The Science Inside

Defence Science and Technology Laboratory

Dstl delivers high-impact science and technology for the UK's defence, security and prosperity.

We deliver world-leading AI and autonomy work in partnership with our customers, user community and specialists from industry, academia and allied nations. We look at everything, from very early research into how machines interact with humans, to applying AI and autonomy to real-world challenges and operational requirements.

In our role, we are able to provide clear advice on the engagement with and use of AI and autonomy in a defence & security environment.

Introduction

This is what we call a Biscuit Book, something you can pick-up and dip into with a tea and biscuit. The Biscuit Book is arranged as a series of easily digestible chunks that cover the individual building blocks of AI and Autonomy, doing so in a manner that provides the essential information without ever being too technical.

We hope you find the Biscuit Book both informative and digestible.

Time for a brew?

What do we mean by AI and Autonomy?

In the first **Biscuit Book: AI, Data Science and (mostly) Machine Learning**, we provided this definition for AI:

Artificial Intelligence

Theories and techniques developed to allow computer systems to perform tasks normally requiring human or biological intelligence.

But this definition is just one of many. And, AI is not really all that intelligent. The AI we know today is very specialised: it is a critical part of a self-driving car, but the AI in that car won't be able to beat you at Scrabble or appreciate music.

Here we define Autonomy:

Autonomy

The characteristic of a system using AI to determine its own course of action by making its own decisions.

And from there we define Autonomous Systems:

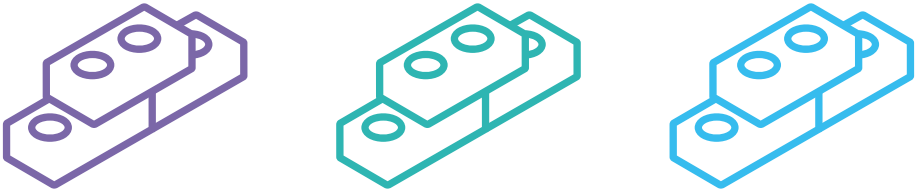
Autonomous Systems

A system containing AI-based components that allow it to exhibit autonomy.

An autonomous system could operate in the physical world (e.g. a self-driving car deciding when to accelerate and brake) or in the virtual world (e.g. an email spam filter deciding which emails to block), or in both.

In truth, it's not a hard and fast line, the key point to make here is that we are talking about systems with AI as a component – the key component perhaps, but only one part of a wider system. In this biscuit book we aim to provide advice and a framework for thinking about AI and Autonomy from a systems perspective.

Building Blocks



So, what do we mean by Building Blocks and what are they? The Building Blocks are a framework we have developed to help break down the complex topic of AI and autonomous systems into more easily accessible chunks. They do not specify how something should be done, but articulate what factors need to be considered.

We have nine Building Blocks that very deliberately provide a high-level perspective. Each Building Block represents something that needs to be considered or actions that need to be taken to ensure the AI and Autonomy is built correctly, does what it's meant to and no more, and does so legally and ethically (yes, it's not just about programming).

Don't think of these building blocks as like breeze blocks, all the same size and the same for each system you work on. A better analogy is Lego™ (other toy bricks are available, we just can't think what they are). Here you have a choice of different sized, shaped and coloured bricks that you select based on what you are trying to build.

The point is that the value that each block adds to a system will depend on the needs of that system and this will vary depending on the maturity of the work, the area of use and other factors. Therefore adequately addressing each Building Block will require some degree of engagement with experts, or reference to supporting materials, to ensure each Building Block is properly understood.

The nine Building Blocks are themselves split into three elements, shown in Figure 1:

- The **Core elements of AI** (**in purple**): Data, Algorithms, Platform and Integration. These are the things that together create AI and hence allow the system to exhibit autonomy.
- The **Critical factors for success** (**in green**): Advantage, Consent and Confidence. These are not part of the system itself but are critical to the successful development and deployment of the system.
- The **Cross-cutting enablers** (**in blue**): Expertise and Enterprise. These are the things that bring everything together and are required to deliver autonomous systems at scale.

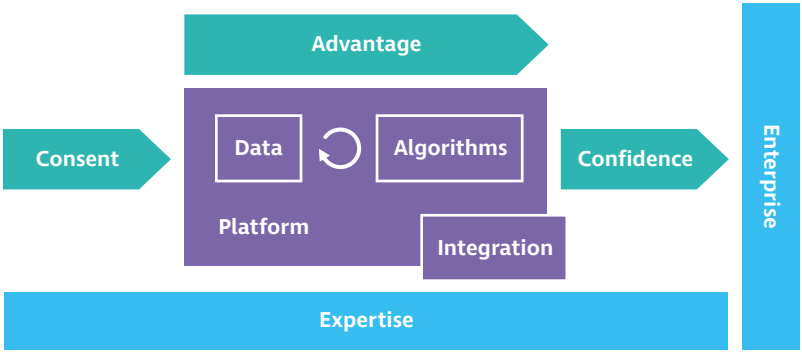


Figure 1 – Building Blocks for AI and Autonomy

Each of the Building Blocks is explained in a little more detail on the following pages (but only a little more to limit the number of brews required).

We also include some questions you should ask to reduce the risk of OSINTOTs (an OSINTOT is what you say right at the end of a project when you realise there is that thing you should have done and your multi-million pound aircraft is now a smouldering heap – "Oh sugar, I never thought of that". Actually in that situation you might not say "sugar".)

Core Elements of AI

Data Element

Block 1 

We must have access to the right data, for both development and operational use of any AI. This could include any source, from our own systems and sensors, to large public datasets, with different implications for privacy and security, and different ownership and usage rights. The one thing to remember about data is that it is more difficult than most people realise. And it's difficult in so many different ways – data likes diversity. So handling data generally takes longer than you think and then some. But, without data you can't do anything – it's that important.

Here are some of the questions you need to ask and things to consider:

Data requirements and availability

- What data do we need to use throughout, from development, training (the AI) and operational use?
- Does that data exist in a usable format?
  Quite frequently the answer is either 'no', or 'yes, but ...'
- If the data isn't right are we able to wrangle it into shape?
- Most importantly, do we understand the data, both structurally and semantically? You'll often find the data is poorly documented and you have to rely on finding someone who really understands it – this in itself can be a challenge.

Data collection

- If the data really isn't available there are really only three options:

  1. **Give up** – Although this is the easiest option, it's not guaranteed to fill the end user with joy. Though if collecting data is very hard or costly, then maybe this is the right answer.
  2. **Collect the data** – An option if you can and a major task in itself, but at least you will have a good understanding of the data at the end. Or, at least you think you will, it might be worth checking that you do.
  3. **Generate synthetic data** – This might be a good option if you really understand the data. It is challenging although there are more and more tools being created to help with this process.

Data labelling

- A lot of AI algorithms require quality labelled training data to learn what to look for (this approach is known as supervised learning – see the Biscuit Book on AI, data science and machine learning for more detail).
- So once you've got your hands on some data, you may need to think about how to label this data.
  Questions to think about:
  - What labels do you want to use?
  - How difficult and time-consuming will this be?
  - Who has the right expertise to be able to label data?

Block 1

Representative and balanced datasets

- Is the data sufficiently representative of all cases where we will want to use the AI or Autonomy? In other words is the data good enough for our purpose.
- You need to think about how data may change over time. Is a training data set that was collected in one context or period of time going to be representative of the data a system might encounter when it is used in the future?
- Do we understand where bias exists in our data, whether this matters and how we can mitigate if so? First thing to recognise is that all data will be biased in some way, so we should understand what the biases are, do they matter and if they do, what can be done about it? Ignoring them is a common tactic (not to be recommended!).

Data ownership and access

- Do we have appropriate arrangements to access and use the data? An amazingly frequent blocker. Even if the data exists and is what you need, are you allowed to get at it? Don't assume you can, or that it won't take time to acquire.

Confidential and private data

- Can we handle sensitive data securely? The essential thing here is having the right methods and secure storage. Again, this can take time to set up.

Algorithms Element

We must have appropriate algorithms (the approach a computer takes) to do what we want to do – the goal if you like. They determine what needs to be done, with what and when, and they must be right for the task. This is easy to state but not always easy to prove. Algorithms must be sufficiently robust, and able to deal with the complexity and constraints of the environment in which they operate – what works well in one environment might be terrible in another – you can't assume anything.

Here are some questions you'll need to ask:

Algorithm selection

- Do we have appropriate algorithms and techniques to solve this problem? Do we understand how the algorithm works and its limitations?
- This could be seen as a trick question: fully understanding how a machine learning algorithm works is near impossible. So really this is about understanding enough about how the algorithm works, and the situations it is likely to be suited for and those it is not.

Block 2 

Implications on, and constraints applied by, other Building Blocks

- What is the impact of algorithm choice on the other Building Blocks, and what constraints do they impose?

  Examples are:

  - Do we understand the trade-offs between algorithm complexity and available compute power: does the compute platform have enough grunt to support the algorithm's demand?
  - What about the trade-off with data: some algorithms may offer high performance but require extensive (and therefore costly and time consuming!) data labelling efforts.

Robustness

- Are the algorithms and techniques able to deal with errors or noise in input data? Are they likely to crash or go wrong for other reasons?

Usage rights and licencing

- Do we understand our legal usage rights where we wish to use algorithms produced by others? This is important because more often than not you will not be developing total solutions but utilising code and algorithms developed by others (including Open Source) – the question is do they permit this?

Platforms Element

The platform comprises the hardware and software where the AI is developed, and where it is then run. These are usually different platforms – it's not a sensible approach to conduct development on live systems! – so we need to think about how to move a useful AI application from development to live.

These platforms could be completely standalone, but more often will be part of an interconnected system. We therefore need to ensure the design of the platform is appropriate, with technical and contractual arrangements to enable timely and cost-effective integration and updates, and with sufficient data storage, compute capability and connectivity where required.

Some questions you should ask are:

Operational platform

- Where are the AI and its data physically hosted; who is providing the hardware platform?
- Do we have the right technical and contractual arrangements to deploy, maintain and upgrade though life?

Development platform

- What development platform will we use?
- Is this accessible to all relevant parties (government, industry and academic partners, international partners, etc.)?
- What is the route from development to deployment? (And, how many of us never think of that until it's too late?)

Block 3

Architectures and standards

- What architecture will be used, and is it open?
- What interfaces will we use?
- Can we re-use system components across the organisation?

Connectivity and compute

- Is there sufficient compute power, in the right places, or can we access it remotely?
- How much data needs to move around the system, and is that achievable?
- Have we considered the merits of AI at the edge (for example a drone might have enough onboard computer power to enable it to act completely independently) versus centralised systems, or novel alternatives?
- Is the connectivity and compute robust and resilient? What happens if the wifi goes down!?

Integration Element

We must consider AI and Autonomy within the context of the wider capability it supports. This means working out how an Autonomous System will interact with other systems (a system of systems!), and potential implications of introducing an Autonomous System – both positive and negative.

Importantly, an Autonomous System will need to interact with people in some way, whether that be the operators or a pedestrian who could otherwise be run over by an overexcited autonomous car. It is therefore critical to understand how people and machines will interact to do the intended job and to do so safely and efficiently.

This means not simply assuming an Autonomous System will replace human operators without impacting the wider system. Equally, where people and machines work together it is important that the person still has a valued role and also does not have to spend significant amounts of time correcting what the machine has just done.

Here are some questions you'll need to ask:

Systems integration

- Are we clear how the AI and Autonomy integrate into the wider Autonomous System?
- And how does that system integrate with other systems (it's rarely completely isolated)?

Block 4 

Ergonomics and interface design

- How will operators interface with the completed system? How well are the needs of people being considered?
- Is it intuitive, easy and clear to the operator how to achieve the desired outcome?
- Is the interface similar to other systems?
- Can we use common components to do this?

Human-autonomy teaming

- How does the Autonomous System work with the user or operator, as part of the team, rather than being just a tool that they need to operate?
- How does the user maintain meaningful control, where necessary?
- Are there any potentially negative implications of introducing autonomy, e.g. could autonomy applied to one role adversely affect a users' ability to perform a different role?
- Will the user get bored? This is important but often overlooked – if the machine does most of the work and leaves only tedious bits to the user, the overall system performance could decline due to the user becoming bored or simply loosing concentration.

Interoperability

- How does our system interoperate with others?
- Does the system need to interact outside of our own enterprise boundary, i.e. other than trained operators using our own systems?
- Will it encounter the public, and will it need to interface with other machines outside of our control?

Critical factors for success

The next three Building Blocks represent critical factors we must understand at the appropriate stages of development and deployment. They aren't part of the system, but are critical to its success.

Advantage

It's a good idea to know why you're doing what you're doing, and in particular what the intended business or operational advantage (benefit) is. This includes understanding the benefits we hope to achieve, the user needs we are trying to satisfy (recognising sometimes we need to experiment to understand these), and the cost implications of our potential solutions across the wider Defence and Security enterprise.

Here are some advantageous (sorry) questions to ask yourself:

Benefit
▪ Are we clear on what we are trying to achieve, and what benefits we expect? It's always a good idea to know this. ▪ Are we confident that the benefits are achievable? Will partially achieving the benefits still be worth doing? ▪ Is this a precursor to gaining more significant benefit in the future? ▪ How will this capability compare with our competitors? Even if we achieve what we set out to, if it's not at least as good as what others are doing, is it worth doing?

Block 5 

User needs

- Do we have a clear understanding of the user needs we are seeking to address?
- Do we have users involved in the project? Here involved means not only turning up to the kick-off meeting but actually being part of the team with a voice that will be listened to.

Cost

- Is there a financial or other cost imposed elsewhere on the organisation, and is the benefit still worthwhile given that cost?
- Are all of those who may be affected involved in the debate?

Protecting advantage

- It's a fact of life that someone will try to disrupt or remove the advantage your system provides. So it's critical to think about how you can stop them and protect your advantage.
- What are the threats to your system? What are the vulnerabilities? How can you protect your system to mitigate these? This might include cyber security, protecting your data and hardening against adversarial AI. You need to think about this for all the Building Blocks!

Intellectual Property and Sovereignty

- If we gain significant advantage from AI and Autonomy, have we considered whether and how we should own or protect the resulting Intellectual Property?
- Should we maintain control over key components?

Consent

Block 6 

We must have consent for the idea and the associated capability. "Consent" is used broadly to include legal and regulatory constraints imposed upon us, as well as satisfying our own policy, ethics and risk appetite, and the willingness of suppliers and partners to support where required.

Here are some questions you should ask yourself:

Legal

- Are there any externally imposed constraints on our capability, such as legal and regulatory frameworks that we need to follow?
- Have we checked the international position as well as domestic?
- What do we need to do to stay within these constraints?
- Is the legal position clear or ambiguous? Do we need to get advice to ensure we comply?
- Is it possible to influence those constraints if we can't operate within them?
- Note that anything involving legal matters will take longer than you can possibly imagine, so factor this in.

Block 6 

Policy and risk appetite

- Is the enterprise (including partners, suppliers and collaborators) likely to be willing to pursue this capability, based on its own internal policies and risk appetite?
- What are the existing policy and risk positions of our organisations?
- Are there international policies to consider?
- What do we need to do to stay within these constraints?
- Is the policy position clear or ambiguous? Do we need to get advice to ensure we comply?
- Is it possible to influence the policy if we can't operate within it?
- Should we try to influence this? For example, what are the risks of not developing the capability?

Ethics

- Fundamentally, should we pursue this capability?
- Have we considered the ethics of doing so, and equally the ethics of not?
- What is our organisation's existing ethical position?
- Does this capability operate within that position?
- Do our ethics align with those of our partners, and will these partners support and engage in our work?
- Are systems fair and equitable?

Confidence

Block 7 

We must have confidence in our AI and Autonomous Systems, and be able to satisfy others of that. "Confidence" is used broadly to include:

- Satisfying regulatory and safety requirements;
- Inspiring trust through assurance, explainability and effective exercising;
- Being aware of the risks through an understanding of threats, vulnerabilities, means of failure and wider resilience.

Relevant questions are:

Assurance

- Will we be able to certify that the system satisfies all relevant regulations, including safety and security standards?
- Will all of the functions that the system performs work reliably, expected and for as long as they need to? The latter is an important point if you have a learning system where the performance could change over time – how do you understand and maintain performance?
- Do we have an understanding of behaviours the system must not have (e.g. harming people – this is generally considered to be a bad thing) and how they can be prevented?
- Do we understand what level of assurance is required?

Block 7 

Trust

- Who needs to trust the system, what do they need to understand and what do you need to provide to obtain this trust?
- This sounds like a simple question but can have many facets – there will be different trust considerations for the direct users, those making decisions based on its outputs, the regulators and the general public.

Explainability

- Do we need to be able to explain why the AI made a particular decision; both at the time, and in retrospect? If so, how can we do this? This is another question that may impact on your algorithm selection: if you really need to know why the system produced a certain output, some types of algorithm will be more suitable than others.

Resilience

- Do we understand the vulnerabilities in the system, and the risks it might introduce to our operations or business? Will the system fail gracefully if it encounters situations beyond its design parameters?

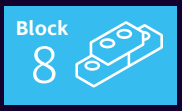
Experimentation

- How suited is the system for experimentation, to build experience and confidence before it is used in a live environment?

Cross-cutting enablers

The final two Building Blocks are cross-cutting enablers, bringing everything together to ensure we are able to deliver a capability that can create the anticipated benefits. These are expressed in terms of Enterprise and Expertise.

Enterprise



No, not the starship, but the wider enterprise that we operate in and within which we are delivering AI and Autonomy: our own organisation, partners, industry, academia – the list goes on. We must ensure this enterprise is prepared for Autonomous Systems, in every aspect from procurement and maintenance, to organisational governance and appropriate training for our people.

Acquisition

- What's the right delivery model and what are the financial and commercial arrangements to support this?

Supply base

- Do we have access to the thriving and diverse supply base for AI and autonomy? Do we know whether we're getting the best capability for what we can afford?

Block 8 

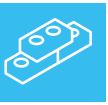
Organisational readiness and governance

- Is it clear where within the organisation a new capability will fit, how it will be governed, what constraints it might operate under, and who will be training users?
- Is your organisation culturally ready to adopt AI and autonomy, or could you encounter resistance?
- Are the right information, infrastructure, logistics and other enablers in place?
- Do we understand the impact it may have on other parts of the organisation?

In-service support

- Do we have a plan for how to maintain and manage the system once it is in use?
- Will the AI learn from experience and be updated, and if so how?
- Is the platform likely to be modified, and if so, how do we keep the Autonomous System in-step with changes?

Expertise

Block 9 

Each of the other Building Blocks must be supported with access to the right expertise through life. This expertise will come from both inside and outside of the business, and possibly from external organisations too, as appropriate to build a diverse team of cross-functional experts.

Within the organisation

- Do we have the right skills and experience across all of the Building Blocks to result in a successful capability?
- Do we have the capability to continually train and retrain our staff as the discipline develops and evolves?

Access to outside expertise

- How are we working with partners and suppliers to access the right skills and experience from wherever they are best sourced?

Diversity

- Is there sufficient diversity (of gender, of ethnicity, of thought) in the teams developing future AI solutions? As well as have the right mix of skills and experience, we need to think about the range of perspectives we need in our multi-disciplinary teams.
- Are we offering the right incentives to encourage others to work with us?

And finally ...

We hope the Building Blocks will excite everyone to focus more attention across the full breadth of the AI and Autonomy challenge, and to think about these issues much earlier in the lifecycle of developing new Autonomous Systems.

By looking across many organisations and projects, using common terms, we expect to be better able to identify the most widespread and significant barriers to adopting AI and Autonomy, and can therefore focus efforts on addressing those areas.

Working with Dstl on AI

Dstl works with its partners and suppliers to deliver research on AI and Autonomy. This research aims to understand how we can responsibly & safely apply these techniques to a wide range of Defence and Security challenges including military decision making, autonomous platforms, computer network defence, sensing, defence logistics, policing and security, streamlining back-office functions and a whole lot more.

To make it simpler to engage with us on AI, Dstl has established AI Lab, a pan-Dstl flagship for AI, Machine Learning and Data Science. We work with suppliers and partners to establish a world-class capability in the application of AI-related technologies to Defence and Security challenges. If you'd like to understand more about Dstl's work on AI and how to engage, you can email:

ai_lab@dstl.gov.uk

[dstl]
The Science Inside

T +44 (0) 1980 950000
E centralenquiries@dstl.gov.uk

For more information about Dstl's work,
visit **www.gov.uk/dstl**

DSTL/PUB126301 v1.0