AI agents inside defence and public safety

What they can read, what they can change, and who decides

IN BRIEF

01 Defence ministries, police forces and border agencies have put AI assistants to work, and most began with paperwork. One defence procurement body describes a chatbot configured “to only draw information from a relevant set of Defence documents”. One home affairs agency says its procurement tool is “shortening processing time by up to 30%”.

02 The record is mixed. A police force switched off its AI assistant after made-up material reached a recommendation. A prosecutor’s office declined police narratives “produced with the assistance of AI”. On where an agent runs, the sector is ahead of most: suppliers offer AI on networks with no outside connection, and agencies have built their own. The report-drafting product for police whose page we read is different: it stores its data in the supplier’s network and does not keep the draft.

03 This paper is about desk work only: the reading, checking and drafting done by staff in policy, logistics, maintenance, procurement, records, training, customs and finance. It is not about operations. It sets out what an agent can read, how it reaches an organisation’s systems, what it may change without a person, and what the rules already say. It ends with seven questions for any supplier, and a place to start.

04 The position throughout is simple. An agent prepares. A named officer or official decides. No agent takes part in any use of force, none judges or identifies a person, and none handles a live emergency. Where the agent runs is settled by the classification of the information. What it may call, as whom, with whose approval and with what record is still open.

IN THIS PAPER

01 Where defence and public safety organisations are today 2
02 The work nobody has packaged 3
03 Three doors, one gate 4
04 What an agent may do 5
05 What the rules already ask 6
06 A worked example: one ministry and one police force 8
07 Seven questions before an agent touches a file 9
08 Where to start 9

PART 01

Where defence and public safety organisations are today

Two
warnings, both from police work

Unchecked
output: the lesson of the first warning

Several
suppliers offer AI on a network with no outside connection

Neither
police product page we read offers on-premises hosting

Defence began with documents. In 2024 the United Kingdom's defence procurement body described "an MoD Cloud hosted, defence data-trained large language model (LLM) chatbot", configured "to only draw information from a relevant set of Defence documents" [5].

Public safety began there too. Singapore's home affairs science agency said its Paperwork tool checks procurement papers by "flagging factual errors, logical inconsistencies and missing information", "shortening processing time by up to 30%" [4].

Two warnings, both from police work. In January 2026 The Register reported that a West Midlands Police recommendation to bar a football club's fans "was partly based on material generated by the AI tool about disruption to a non-existent match" [8]. The acting Chief Constable said: "On my first day in office, I took the immediate step to turn off access to Microsoft Copilot within the force" [8]. In 2024 a prosecutor's office in Washington State wrote: "For now, our office has made the decision not to accept any police narratives that were produced with the assistance of AI" [9]. The first case was intelligence work, outside this paper. Its lesson is not: the output went unchecked.

Where an agent runs is largely settled, and several suppliers offer it. Google describes "Our GDC air-gapped product, which is now authorized for US Government Secret and Top Secret missions, and on which Gemini is available" [1]. Microsoft announced a model "now available in our Secret and Top Secret air-gapped clouds" [2]. Scale AI says: "Donovan is available on classified and air gapped networks" [3]. Singapore's agency built its own, an "on-premise sovereign AI infrastructure", so that "the Home Team's data never leaves the agency's control" [4]. Katonic is one of several that run this way.

The police report products are the supplier's service. Axon says: "All Draft One data, including body-worn camera transcripts and report drafts, is securely stored and managed within the Axon Network" [6]. Neither police product page we read offers on-premises hosting [6][7].

Two kinds of work in one organisation. Operations draw the attention, and the hardest rules. Behind them sits the work this paper is about: the record checked, the contract summarised, the disclosure prepared. An organisation needs a clear line between the two.

What the published accounts leave out. The pages we read say where the model runs.

Few say which tools an agent may call, as whom, with whose approval, and what record is kept.

PART 02

The work nobody has packaged

None

was found for the desk jobs listed on the right of the table

Two

rows on the left sit beside live emergency work, which this paper leaves out

We listed the desk jobs in a defence or public safety organisation and looked for each in the template libraries of the agent and workflow platforms, and on suppliers' own pages. What exists gathers around the police report, the control room and the general assistant.

WHERE PRODUCTS OR PREBUILT AGENTS EXIST	WHERE NONE WAS FOUND
A report narrative drafted from body-worn camera audio	**A completed maintenance record checked against the technical publication**
Report fields suggested from field recordings	**The spares a planned task needs checked against stock and open orders**
Video and audio marked for redaction	**A disclosure request prepared, with each possible exemption listed**
Staff questions answered from a set of defence documents	**A case file checked against the file standard before it goes forward**
Procurement papers checked for errors and gaps	**A unit's training records set against the requirement**
A general assistant on a classified network	**A goods declaration set beside its commercial documents**
Non-emergency calls answered	**An allowance claim checked against the regulations**
A caller's location suggested to the call-taker	**An exercise or incident record assembled into an after-action report**

Two findings matter more than the list. First, the report-drafting product keeps a person in the loop and loses the draft. Axon says officers "must review their reports, fill in the blanks, and sign off on the accuracy of each report before submitting", and that each use is logged "to support transparency without storing the draft itself" [6]. The Electronic Frontier Foundation's reading: "there's no record showing whether the culprit was the officer or the AI" [10].

Second, the last two rows on the left sit beside live emergency work. Motorola Solutions says its tool leaves "the final decision to the telecommunicator" [7]. This paper leaves that work out.

The gaps sit where the reading is slow and the decision already belongs to a named officer or official. That makes them the safer place to begin.

PART 03

Three doors, one gate

An agent is only as useful as what it can reach. In a defence or public safety organisation that means three kinds of thing.

Documents.

Policies, doctrine, standing orders, technical publications, procurement rules, training material. These are indexed once as knowledge sources, with each reader's permissions and markings kept, so an agent answers with the paragraph it relied on and says when the documents are silent.

Systems of record.

Stores and spares, work orders, contracts, personnel and training records, the records index. Most are reached through an interface the organisation already runs. The organisation gives the platform the description of that interface, chooses which operations become tools, and decides which of them need approval.

Logistics and readiness data for analysis.

Stock, availability, demands, delays and training currency, held in the data warehouse or a reporting database, and read only. Dispatch, command and weapon systems, and intelligence holdings on people, are not one of the doors.

All three pass one gate: every tool call is checked against the organisation's policy, which can allow it, record it, hold it for a person, or block it.

Where the organisation's interface checks the caller, the agent acts as the person using it and cannot claim another.

SIX WAYS INTO AN ORGANISATION'S SYSTEMS

	ROUTE	WHAT IT REACHES	WHAT TO KNOW
01	**A file the user uploads**	An officer's notes, a supplier's bid, a maintenance form, a disclosure request	Nothing is connected. The safest start
02	**A knowledge source**	Policies, doctrine, standing orders, technical publications, procurement rules	Permissions and markings are kept per reader
03	**An imported description of an interface**	Stores, work orders, contracts, training records, the records index	You choose the operations. Changes wait for approval
04	**A reporting database**	Stock, availability, demands and delays, training currency	Read only
05	**A server from the tool catalogue, run in your cluster**	Databases, document conversion	It runs inside; check what each tool does
06	**Public sources, where the network allows**	Laws, published guidance, tender notices, sanctions lists	Reads only. On a closed network they come in as files

PART 03 · CONTINUED

27
public interface descriptions read, at the edge of this work

None
refused by the importer

123
operations in those descriptions

60
of those operations are sent as changes

What is published. We read 27 public interface descriptions at the edge of this work: public procurement, customs and border, tenders and sanctions lists. The importer refused none. Together they carry 123 operations, of which 63 read and 60 are sent as changes [11].

What that exercise taught us. The open interfaces are built to do things, not only to read. Nearly half of the operations are sent as changes.

What is not published. The operational systems publish no public description: police records, dispatch, digital evidence, defence logistics and maintenance. The organisation supplies its own description, inside its own network.

Classification and the gate. The safe design keeps the agent on the desk side. It reads nothing its user is not cleared to read, and it reaches no system that directs a unit or a weapon.

So choose the operations at import, start with the ones that read, and add nothing that files, orders or releases until the approval step has been seen working.

PART 04

What an agent may do

One rule runs through every agent in this paper: it prepares, and a named officer or official decides.

01	**It reads** what the person using it is cleared to read, and no more.
02	**It drafts** the briefing note, the check of a record, the summary of a contract, the report from an officer's notes, with each statement tied to its source.
03	**It lists** what is missing, what does not match, what falls due and when.
04	**It stops** before anything that changes a record or leaves the organisation. A tool that changes something waits for approval.
05	**It decides nothing about a person.** It does not identify, profile, score or predict anyone, does not recognise faces, and does not infer anyone's emotions.

PART 04 · CONTINUED

Four lines are drawn tighter than the rest.

People and decisions about them.

An agent may check a claim against the regulations, or list the qualifications due in a unit. It collects no intelligence on people. It takes no part in a decision to arrest, charge, detain, search, or admit a person at a border.

Reports and evidence.

An agent may turn an officer's own notes into a report in the force's format. The draft is marked as written with an agent, and the notes are kept beside it. The agent adds no fact the notes do not hold. The officer reads, corrects and signs the report, and owns every word. An agent does not weigh how reliable a piece of evidence is.

Operations and emergencies.

No agent takes part in targeting or in any use of force, and none dispatches a unit. This paper leaves live emergency work out: no agent takes an emergency call, triages it or dispatches to it. Emergency services appear here for their office work only: policy, training, records, after-action reports.

Classified information and where the agent runs.

The classification of the information settles where the agent and its model run, up to a network with no outside connection. An agent and its model stay on the network they are installed on. Moving anything between networks stays the organisation's own controlled process. Its security authority decides what may run where.

PART 05

What the rules already ask

This part needs a lawyer's read before it is relied on. It reports what the texts say; it is not advice.

What is prohibited

Four practices the European Union has banned since 2 February 2025. The AI Act prohibits the use of an AI system "for making risk assessments of natural persons in order to assess or predict the risk of a natural person committing a criminal offence, based solely on the profiling of a natural person". It prohibits systems "that create or expand facial recognition databases through the untargeted scraping of facial images from the internet or CCTV footage", and systems "to infer emotions of a natural person in the areas of workplace and education institutions" [12].

A fourth is banned with narrow exceptions: "the use of 'real-time' remote biometric identification systems in publicly accessible spaces for the purposes of law enforcement" [12]. No agent in this paper does any of the four.

Law enforcement and borders

The high-risk list names both. For law enforcement it includes systems "to evaluate the reliability of evidence in the course of the investigation or prosecution of criminal offences". For borders it includes systems "to assess a risk, including a security risk, a risk of irregular migration, or a health risk, posed by a natural person who intends to enter or who has entered" a member state [12]. Drafting a report from notes and checking a goods declaration are not on the list as we read it. Whether a given tool falls under it is for counsel.

Emergency calls are on the list too. It names systems "intended to evaluate and classify emergency calls by natural persons or to be used to dispatch" emergency services [12].

The dates have moved. Regulation (EU) 2026/1744 says the date of application of the high-risk duties "is set to 2 December 2027 for AI systems classified as high-risk pursuant to Article 6(2) and Annex III" [13]. All three entries sit in Annex III.

A court has tested a police technology. In Bridges v South Wales Police the Court of Appeal declared that a force's use of live facial recognition "was not in accordance with the law for the purposes of Article 8(2)" [16]. Facial recognition is outside this paper. The case shows that a court asks first for a clear legal basis.

PART 05 · CONTINUED

Defence

The AI Act leaves exclusively military systems out. Article 2(3): "This Regulation does not apply to AI systems where and in so far they are placed on the market, put into service, or used with or without modification exclusively for military, defence or national security purposes" [12].

What that means, and does not. The word that matters is "exclusively". The text does not say that everything a defence ministry does is outside it. Whether a ministry's office work with a general system is inside or outside is for counsel.

Defence writes its own rules, and they reach the office. The United Kingdom's Ministry of Defence says its ethical principles "are applicable across the full spectrum of use cases for AI in Defence, from battlespace to back office", and that "Human responsibility and accountability cannot be removed" [15].

Reports, records and disclosure

One American state requires the report to say so. Utah's law of 2025 says a police report "created wholly or partially by using generative artificial intelligence" shall "contain within the report or record a disclaimer that the report or record contains content generated by artificial intelligence", and a certification by its author that the author has read and reviewed it for accuracy [17].

The public must be told. Under the AI Act a system "available for the public to report a criminal offence" must say that it is an AI system [12].

MARKET	WHAT WE READ	WHAT IT ASKS
European Union	AI Act, Article 14(1)	High-risk systems are built so "that they can be effectively overseen by natural persons during the period in which they are in use" [12]
European Union	Directive (EU) 2016/680, Article 11(1)	A decision based solely on automated processing is prohibited unless a law authorises it, with "at least the right to obtain human intervention on the part of the controller" [14]
United Kingdom	JSP 936, part 1	A directive for defence, described as "the principal policy framework governing the safe and responsible adoption of AI in MOD" [15]
United States, Utah	S.B. 180, section 53-25-602(1)(a)	"A law enforcement agency shall have a policy concerning the use of generative artificial intelligence" [17]

We did not open United Kingdom data protection law, United States federal rules, NATO's principles, or the rules in the Gulf, India, Singapore or Australia. An organisation there should ask its own counsel.

PART 05 · CONTINUED

What the rules have in common

01	**Some uses are closed,** whoever asks: profiling to predict crime, scraping faces, inferring emotions at work.
02	**A person decides** anything that affects a person, and can be shown to have decided.
03	**A report says how it was written,** and its author answers for it.
04	**Sensitive records need protecting,** which is easier where the organisation controls the system.

PART 06

A worked example: one ministry and one police force

Take a made-up defence ministry with a support organisation of about 14,000 staff, and a made-up police force of about 6,000. Nothing here is a customer's story, and every figure is illustrative; it is section 4 put to work.

01

In the ministry's policy branch, a staff officer asks what the standing orders say about a travel allowance. The agent answers with the paragraph cited and says where the orders are silent. The policy owner rules on that point.

02

In the maintenance wing, the agent reads 220 completed maintenance records against the technical publication and lists 17 with a missing signature or a value outside limits. The authorising engineer accepts or rejects every record.

03

In procurement, nine bids arrive for a vehicle spares contract. The agent checks each against the 34 mandatory requirements and lists what is missing. The panel evaluates and decides.

04

In the force's disclosure unit, a request covers 1,400 pages. The agent lists 96 passages that may fall under an exemption. The disclosure officer decides every one and releases the response.

05

At a police station, an officer uploads her own notes of a burglary. The agent drafts the report in the force's format, marked as written with an agent, with the notes kept beside it. She corrects two lines and signs it.

06

In the training office, the agent lists 58 qualifications due or expired in one unit. The unit training officer decides what is booked.

Every one of these reads documents or records the organisation already has, changes no record by itself, decides nothing about a person, and leaves a record of who asked and what was shown.

PART 07

Seven questions before an agent touches a file

01 **Where does it run?**
On our infrastructure, on our classified network, or in the supplier's service? Where is the model? Can it run with no connection to the internet?

02 **What can it read?**
Does it act as the person using it, with their permissions and markings, or as one all-seeing account?

03 **What can it change?**
Which tools change a record, place an order or release a document, and does each wait for a person?

04 **What will it never do?**
Show us that it cannot be pointed at a person: no identifying, no scoring, no predicting.

05 **Who decides, and who signs?**
Show us the step where a named officer approves, and how a drafted report is marked and its notes kept.

06 **What is recorded?**
Who asked, which tool ran, what was decided, and for how long is that kept?

07 **Can we stop it?**
One agent, one tool, or all of them, at once, by us.

PART 08

Where to start

Pick work where reading is slow, the decision stays with a person, and no operation waits on the agent.

Policy and doctrine questions and answers
Staff ask what the organisation's own policies and orders say and get the answer with the paragraph cited. No system of record is touched.

Maintenance record check
The agent lists missing entries and signatures on a completed record. The authorising engineer accepts or rejects it.

Disclosure request preparation
The agent gathers the records and lists possible exemptions. The disclosure officer decides and releases.

One unit, one process, unclassified or invented data first. Then beside the manual process, measured against it.

About Katonic AI

Katonic AI builds the platform that organisations use to build, run and govern AI agents on their own infrastructure: in their own data centre, in a private or sovereign cloud, or with no connection to the internet at all. The platform is in production with governments, banks, telecom operators and utilities. Katonic is certified to ISO 27001.

KATONIC.AI/SOLUTIONS/DEFENCE

SOURCES

Read on 22 November 2026. Figures are the publishers' own and were not audited by us.

01 Google Cloud, "Bringing Gemini and Google Agentspace to you on-premises", blog post, 9 April 2025. cloud.google.com

02 Microsoft, "Announcing GPT-5.2 Availability in Azure for U.S. Government Secret and Top Secret Clouds", Azure Government blog, 15 January 2026. devblogs.microsoft.com

03 Scale AI, Donovan product page. scale.com

04 HTX (Singapore), "Why sovereign AI is key to scaling public safety impact", 8 June 2026; and "HTX launches Home Team AI Suite to boost AI adoption", 29 January 2026. htx.gov.sg

05 Defence Equipment & Support (United Kingdom), "DE&S exploit AI to drive efficiencies in cyber and security compliance", 8 February 2024. des.mod.uk

06 Axon, Draft One product page. axon.com

07 Motorola Solutions, Assist product page. motorolasolutions.com

08 SA Mathieson, The Register, report of West Midlands Police suspending its AI assistant, 28 January 2026. theregister.com

09 Electronic Frontier Foundation, note of 17 October 2024 that reproduces passages of the King County prosecutor's memo. eff.org

10 Electronic Frontier Foundation, "EFF Investigation: AI Product for Police Reports is Designed to Hinder Audits", press release, 10 July 2025. eff.org

11 Template libraries and suppliers' pages, searched 22 November 2026; and 27 public interface descriptions fetched the same day and run through the platform's importer (Katonic research).

12 Regulation (EU) 2024/1689 (AI Act), Articles 2(3), 5(1)(d), (e), (f) and (h), 14(1), 50(1) and 113(a); Annex III points 5(d), 6(c) and 7(b).

13 Regulation (EU) 2026/1744, 8 July 2026, the recital on the dates of application.

14 Directive (EU) 2016/680, Article 11(1).

15 Ministry of Defence (United Kingdom), "Ambitious, safe, responsible: our approach to the delivery of AI-enabled capability in Defence", 15 June 2022; and the JSP 936 (part 1: directive) page, 13 November 2024. gov.uk

16 R (Bridges) v Chief Constable of South Wales Police [2020] EWCA Civ 1058, 11 August 2020, the declarations. caselaw.nationalarchives.gov.uk

17 Utah S.B. 180 (2025), enrolled copy, section 53-25-602; in effect 7 May 2025. le.utah.gov