PUTTING SCIENCE INTO PRACTICE · SCIENCE TO THE STATION 2026

Human-AI teaming in fire service operations

Evaluating AI decision support before it reaches your fireground

Scott Ramey, MScHQ

PhD Candidate, Systems Design Engineering, University of Waterloo
Founder, The Human Factor (thehumanfactor.ca)
Former Assistant Fire Chief / Deputy Paramedic Chief

Slides, the safety checklist and every source

thehumanfactor.ca/s2s2026-practice

Kansas City · Tuesday, November 3 and Wednesday, November 4 · 2:45 to 3:45 pm · Trianon E

BEFORE WE START

My interests and declarations

DECLARED

The Human Factor

My long-standing consulting agency.

HazReady

Software for emergency decision support, personally developed for my research.

DRDC CSSP funding

$2M over three years. Four universities, a municipality, industry partners.

WHAT THAT MEANS FOR THIS HOUR

- ✓ No product or service is offered in this session.
- 📖 The evidence comes from published research and my PhD work.
- 🔍 The screens are the HazReady research prototype on test data, there to show a principle.

BEFORE WE START

Why I do this

Tech nerd

I build the tools I study. That is a bias to declare.

Firefighter and paramedic

flight paramedic, emergency manager, assistant fire chief, deputy paramedic chief.

Air ambulance helicopter crash

Colleagues lost. Human Factors and Systems Safety.

Fix the system around the person. Not the person.

THE HOUR

One idea, seven parts

AI can make a fireground safer or more dangerous. Which one depends on whether it is designed around the people who use it, tested the way we test everything else, and the judgment stays with the officer.

You'll leave with four measures, a two-hour pilot, a way to read its results, six questions for any vendor, and a safety checklist your department can run on Monday.

1
What AI is
7 min

2
People and AI
10 min

3
What the research shows
6 min

4
Safer, or not
4 min

5
From research to tools
11 min

6
Evaluate before you adopt
10 min

7
Try it
5 min

1 · WHAT AI IS

The organization owns what its AI says

AIRLINE CHAT ASSISTANT · NOVEMBER 2022

My grandmother just died. Is there a bereavement fare?

Yes. Book now, and apply for the bereavement rate within 90 days of travel.

The real policy allowed no refund after travel. The chatbot made that up.

In effect, the tribunal said, the airline was arguing its chatbot was “a separate legal entity that is responsible for its own actions.”

“A remarkable submission.”

The tribunal held the airline responsible and ordered it to pay.

Whoever deploys the AI owns what it says: a department, a vendor, a dispatch centre.

Moffatt v. Air Canada, 2024 BCCRT 149. Chat wording paraphrased from the decision.

1 · WHAT AI IS

A 70-year-old idea, a 3-year-old tool

1955

The Dartmouth proposal names “artificial intelligence”

1966

ELIZA: a few hundred lines of pattern matching, and people confided in it

2017

The transformer: the architecture behind today’s chatbots

2022

ChatGPT, November 30. An estimated 100 million monthly users two months later

What changed in 2022 wasn’t the idea. Everyone got it at once, free, in plain English.

McCarthy et al. 1955 · Weizenbaum 1966 · Vaswani et al. 2017 · OpenAI 2022; UBS estimate via Reuters, Feb 2023

1 · WHAT AI IS

“AI” is four different things, and each fails differently



Recognize

Smoke detection cameras. Thermal imagery from drones.

Fails by misidentifying



Predict

Fire spread. Where to post units. Which buildings are at risk.

Fails by learning yesterday’s patterns



Generate

Drafting the incident report. Summarizing an SOP.

Fails by making things up, fluently



Triage

Sorting 911 calls when the phones light up.

Fails by deciding who waits

First question for any vendor: which kind is this?

Ramey, Masoumi-Ravandi, Rehman and Cao, systematic review (in preparation): systems ranged from triage to forecasting to drone imagery; few were fully operational.

1 · WHAT AI IS

What a chatbot actually does

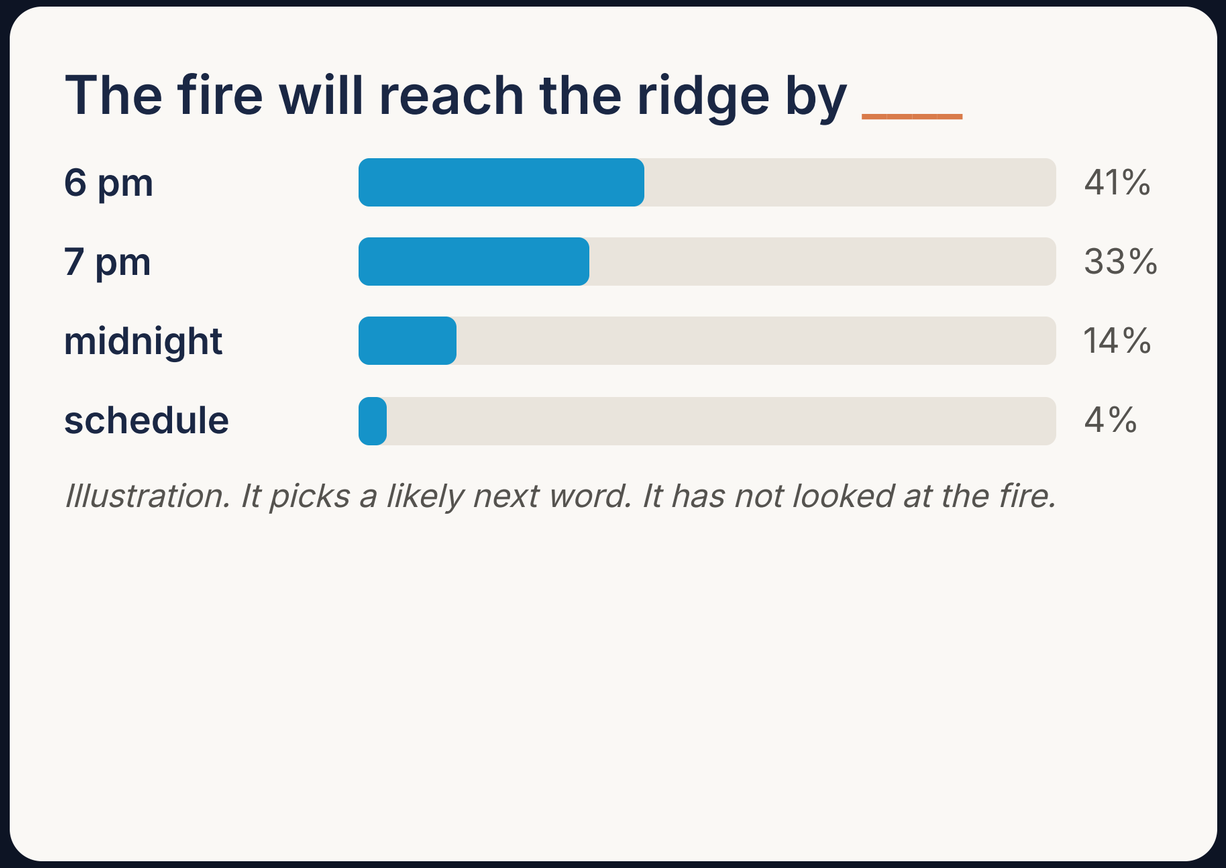


Illustration. It picks a likely next word. It has not looked at the fire.

Generative AI models “generate content that is statistically likely in response to a prompt, for example, by predicting the next word in a sentence.”

Government of Canada, Guide on the use of generative AI

Very sophisticated autocomplete.

No built-in sense of true or false. Not a search engine. It doesn’t look anything up unless it’s connected to something that does.

Treasury Board of Canada Secretariat, Guide on the use of generative artificial intelligence (2026)

1 · WHAT AI IS

Why it sounds so sure

Fluent is not accurate

It writes the wrong answer in the same confident voice as the right one. The industry word is “hallucination.”

Tuned to please

Assistants “consistently exhibit sycophancy.” In 2025 one update was withdrawn for being “overly flattering or agreeable.”

Professionals got caught

Lawyers who cited invented cases: sanctioned in New York, 2023; ordered to pay costs in Vancouver, 2024.

The tone that feels most helpful is the one to be most careful with.

TBS Guide on generative AI · Sharma et al. 2024, ICLR · OpenAI, April 29 2025 · Mata v. Avianca (S.D.N.Y. 2023) · Zhang v. Chen, 2024 BCSC 285

1 · WHAT AI IS

Where the model runs decides who sees your data

Public cloud assistant

The chat app on everyone’s phone.

Most capable. Your prompt leaves the building, on someone else’s terms.

Air-gapped local model

Open-weight model, your own hardware, no internet.

Nothing leaves. Less capable, and every update, test and failure is yours.

Domain-tuned local model

Your SOPs, pre-plans and ERG, retrieved or trained in.

Answers in your context, with sources. Still invents. Still needs the checks.

SHADOW IT

Someone in your department is already doing this, on a personal phone, with real world implications.

TBS Guide on the use of generative AI · Bloomberg, May 2023 (Samsung) · Microsoft and LinkedIn, Work Trend Index 2024

2 · PEOPLE AND AI

Human factors in one slide, Safety Differently.

Design the system around the people who use it, under the conditions they really work in.

> The more we automate, the more we depend on people for the rare, hard cases.
>
> Bainbridge, “Ironies of automation,” 1983 · Endsley, “Ironies of artificial intelligence,” 2023

THE LENS, ONE IDEA PER SLIDE

- The interface shapes the decision
- Trust: calibrated, not maximized
- Why we’re built to believe it, and what stress does
- Automation bias
- A human in the loop is not enough
- Built elsewhere doesn’t fit automatically

Bainbridge 1983, Automatica · Endsley 2023, Ergonomics · Ramey et al., systematic review: AI succeeds only with human-centred design, co-development and real evaluation

The interface shapes the decision

ALERT MENU · HAWAII, JANUARY 13 2018
Amber Alert (CAE) – Kauai County Only
Tsunami Warning (CEM) – STATE ONLY
PACOM (CDW) – STATE ONLY
DRILL – PACOM (CDW) – STATE ONLY

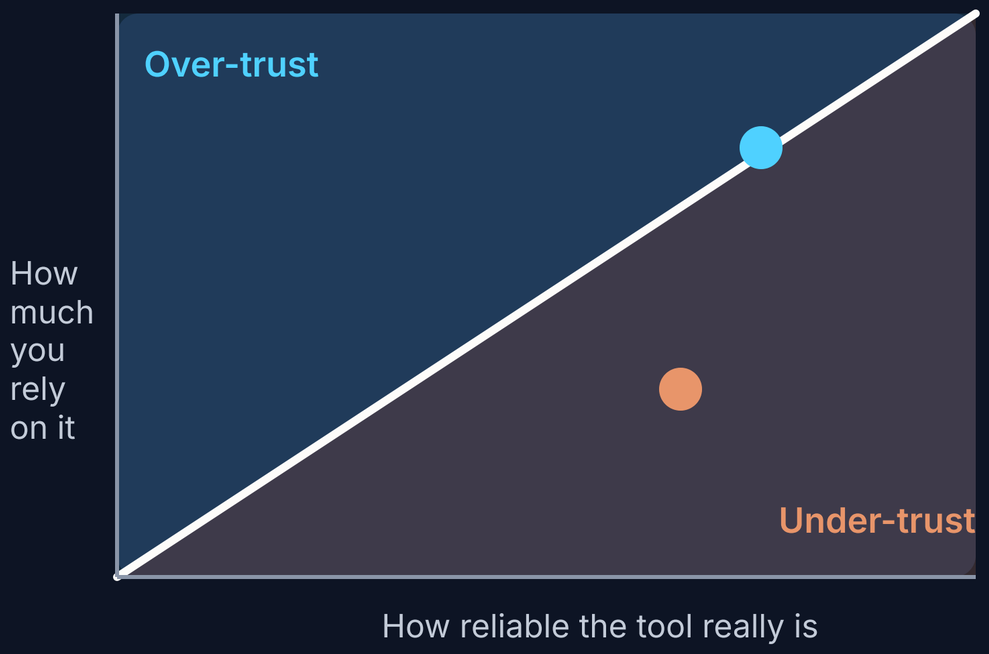
Drill and live alert side by side, behind the same prompt. No plan to correct it: 38 minutes.

- More data on screen “may easily result in harming” awareness and performance
- Show the few things that matter
- Too many alerts become noise

Situational awareness is seeing what matters, understanding it, and projecting what happens next. The screen can build that, or bury it.

FCC 2018 report on the Hawaii false alert (menu wording illustrative) · Parush et al. 2020 · Bai et al. 2025 · Endsley 1995 · Ramey et al., systematic review

Trust: calibrated, not maximized



Rely on a tool exactly as much as it deserves. No more, no less.

Trust follows performance, and people drop an algorithm faster than a person after the same mistake. In emergency services, I think over-trust is the bigger danger.

In a survey of 99 firefighters, rescue pressure changed how capable they judged the AI, and so how much they trusted it.

Pressure moves trust, exactly when checking matters most.

Hancock et al. 2011 · Dietvorst et al. 2015 · Zou et al. 2026, Ergonomics · Ramey et al., systematic review; Ramey and Rehman, scoping review (draft)

Why we’re built to believe it

Fluency

What is easy to read feels true. AI text is the most fluent text you will ever read.

The machine heuristic

“The computer says” feels objective. People don’t.

Confirmation

We accept what agrees with us. The tool is tuned to agree.

Anchoring

The first plan on the table pulls every plan after it. The AI draft is always first.

And the reverse: after seeing it make mistakes, people abandon a tool that still beats them.

Algorithm aversion. Both errors come from the same place: we judge the tool by how it feels, not by its record.

Alter and Oppenheimer 2009 · Sundar 2008 · Nickerson 1998 · Tversky and Kahneman 1974 · Dietvorst, Simmons and Massey 2015

Under stress, the shortcuts get stronger

CALM



Eight cues in view, including the one that says “how sure”.

UNDER STRESS



Two cues. The loudest. The “how sure” line is the first to go.

- **Attention narrows.** Peripheral cues drop out first.
- **Experts take the first workable option.** A confident AI suggestion becomes the first option.
- **Stress closes the search early.** Fewer options considered.
- **Automation bias peaks under workload.** Exactly when the tool is most used.

Under pressure you will look at fewer options than you think. The tool will sound more certain than it is. Design the check for the moment you are weakest.

Easterbrook 1959 · Klein 1989, 1998 · Keinan 1987 · Staal 2004 (NASA) · Parasuraman and Manzey 2010 · Zou et al. 2026

2 · PEOPLE AND AI

Automation bias

People accept automated advice when it’s wrong, and miss what it misses, most of all under workload and time pressure.

> It “cannot be prevented by training or instructions, and can affect decision making in individuals as well as in teams.”
>
> Parasuraman and Manzey, 2010

What helps: making a named person clearly accountable for the decision.

“

What happens if the system provides an incorrect recommendation and the provider follows it? Who is accountable for any resulting medical errors?

The researchers, after interviewing paramedics (Bai et al., 2025)

Parasuraman and Manzey 2010, Human Factors · Goddard, Roudsari and Wyatt 2012, JAMIA · Bai et al. 2025 · Ramey et al., systematic review

2 · PEOPLE AND AI

A human in the loop is necessary. It is not sufficient.

41 OVERSIGHT POLICIES REVIEWED

“People are unable to perform the desired oversight functions.”

Green, 2022

319 WORKERS, 936 WORK EXAMPLES

“Higher confidence in GenAI is associated with less critical thinking.”

Lee et al., 2025

What works: a check built into the task, not a signature at the end

- Nothing is final until a person has acknowledged or overridden each recommendation.
- Check the AI-written report for errors, or watch the footage first.
- Claims that fail an evidence check go to a person, not to the screen.

Green 2022 · Lee et al. 2025, CHI · Spearing et al. 2026 · Tan et al. 2026 · both reviews

2 · PEOPLE AND AI

Built elsewhere doesn’t fit automatically

Hospital data, hospital workflow

Emergency department triage follows protocols. Fireground command is recognition and replanning on the fly.

Double work, extra hands

“Avoiding double documentation is essential.” Gloves, a facepiece and a radio in each hand. Almost no studies test AI with PPE on.

Experts and novices

Across 12 emergency departments: “excelling with trained users but requiring targeted refinements for novices.”

Ask what data it learned from, and whose workflow it was built around.

Klein et al. 1986 · Bai et al. 2025 · Pinero de Plaza et al. 2025, Int J Med Inform · Ramey et al., systematic review (in preparation)

3 · WHAT THE RESEARCH SHOWS

I went looking for the evidence

1,007
records, six databases, 2000 to June 2025

32
studies of responders using AI decision support

Few
in the fire service. The largest group, 13, came from paramedicine and emergency departments.

Trust, usability and fit with the work decided whether AI got used.

Ramey, Masoumi-Ravandi, Rehman and Cao (in preparation). Humans, algorithms, and alarms: a systematic review. University of Waterloo.

3 · WHAT THE RESEARCH SHOWS

Everyone talks about trust. Almost nobody measures it.

8 / 32
named a theory

7 / 32
used any structured measure

0
used a standard, validated instrument

So the evidence behind most AI tools you'll be shown is a demo and a satisfaction survey. No controlled studies with and without AI. Few field trials. No long-term studies.

The good news: you can do better yourselves. Section 6 shows how.

Ramey et al., systematic review (in preparation), Figure 3

3 · WHAT THE RESEARCH SHOWS

My research: done, under way, next

How do incident commanders calibrate trust in, and appropriately rely on, AI decision support under uncertainty?

DONE

Systematic review

32 studies. Written.

Scoping review: governing AI

42 sources, none Canadian. Drafted.

UNDER WAY

Study 1: how commanders decide

Hazmat incident commanders, the first 30 minutes, before any AI. Ethics approved.

NEXT

Study 2: do they catch an AI error?

Simulation, 36 participants. Workload, awareness, trust, overrides. Ethics in review.

Study 3: trust over time

Repeated use over 12 weeks. Ethics in review.

What responders accept

FIREFIGHTERS ON DATA ABOUT THEMSELVES

- Never punitive
- De-identified
- Restricted access
- Voluntary
- They can see their own data

Where US law said little, some unions bargained data limits into their contracts. Protections varied.

AI review of body camera footage felt less fair to officers when they had no control over recording and review.

Acceptance is about control and purpose, not technology.

For a department, that is the design brief: decide control and purpose with the crews, before go-live.

Topazian et al. 2025, JOEM · Topazian et al. 2026, New Solutions · Adams 2025, Journal of Criminal Justice · Ramey and Rehman, scoping review (draft)

4 · SAFER, OR NOT

Better decisions or worse: the four things AI moves

	BETTER, WHEN IT IS DESIGNED FOR THE WORK	WORSE, WHEN IT IS NOT
Situational awareness	It carries the projection: where the plume goes, what changed since you last looked.	Twelve layers and a stale feed shown as current. Attention tunnels on the screen, not the fire.
Workload	It pre-computes the lookups: distances, guide pages, the first draft. Less to hold in your head.	Double entry, an alert a minute, and a tablet to feed with gloves on.
Trust	An honest “how sure” and a visible record: you rely on it as much as it has earned.	Pressure moves trust, and one visible failure switches the tool off for good.
Bias	A named person owns the decision, and the check is built into the task.	The AI’s draft is the first plan on the table. At peak workload, you take it.

Same tool, both columns. The design and the department decide which column your crews get.

Endsley 1995 · Parush et al. 2020 · Bai et al. 2025 · Hancock et al. 2011 · Dietvorst et al. 2015 · Zou et al. 2026 · Parasuraman and Manzey 2010 · Goddard et al. 2012 · Ramey et al., systematic review

4 · SAFER, OR NOT

Where AI may already be on your calls

911 and dispatch
Call triage, translation, transcribed calls.

Structure fire
Accountability boards. A drafted action plan.

Hazmat
Chemical lookup. The plume on the map.

Wildland
Satellite hot spots. Drones that flag people.

Medical
The AED’s shock advice since the 1990s. ePCR narratives.

Vehicle collision
Crash notification from the car. Extrication guides.

Rehab and wearables
Fatigue and heat strain estimated from heart rate.

After the call
Reports drafted. Radio traffic transcribed.

The cost of a wrong output here is paid by a firefighter, not by the vendor.

Endsley 2023 · Parasuraman and Manzey 2010 · Topazian et al. 2025 · Ramey et al., systematic review · Ramey and Rehman, scoping review (draft)

5 · FROM RESEARCH TO TOOLS

Where it leads: four things fire services can use

1 Design guidelines

What AI on the fireground has to do, and must never do: confidence up top, “what changed,” one-tap override, no stale data shown as current.

2 An evaluation toolkit

Four measures and a two-hour pilot any department can run before it buys. You get the first version today.

3 A governance template

Who may override, who is accountable, and the rules for data about firefighters: never punitive, de-identified, voluntary.

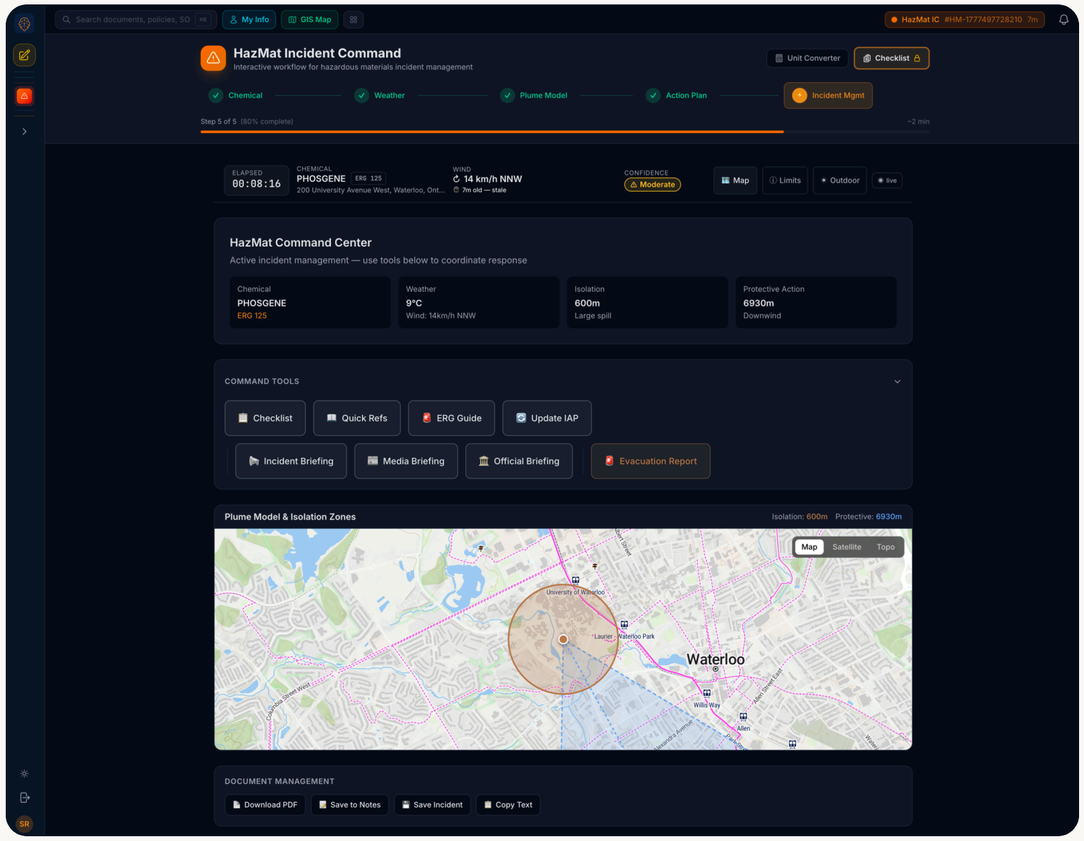
4 A decision-support tool built with commanders

Shaped by Study 1, tested in Studies 2 and 3. The next four slides show what the principles look like on its screens.

Planned outputs of the research program. Screens that follow: research prototype, test data, unbranded.

5 · FROM RESEARCH TO TOOLS

Situational Awareness: Data is not awareness



One row, four things

Chemical, weather, isolation, protective action. Everything else is a tap away.

“Moderate confidence”

Up top, in colour, where narrowed attention still finds it.

“Wind: 7 min old, stale”

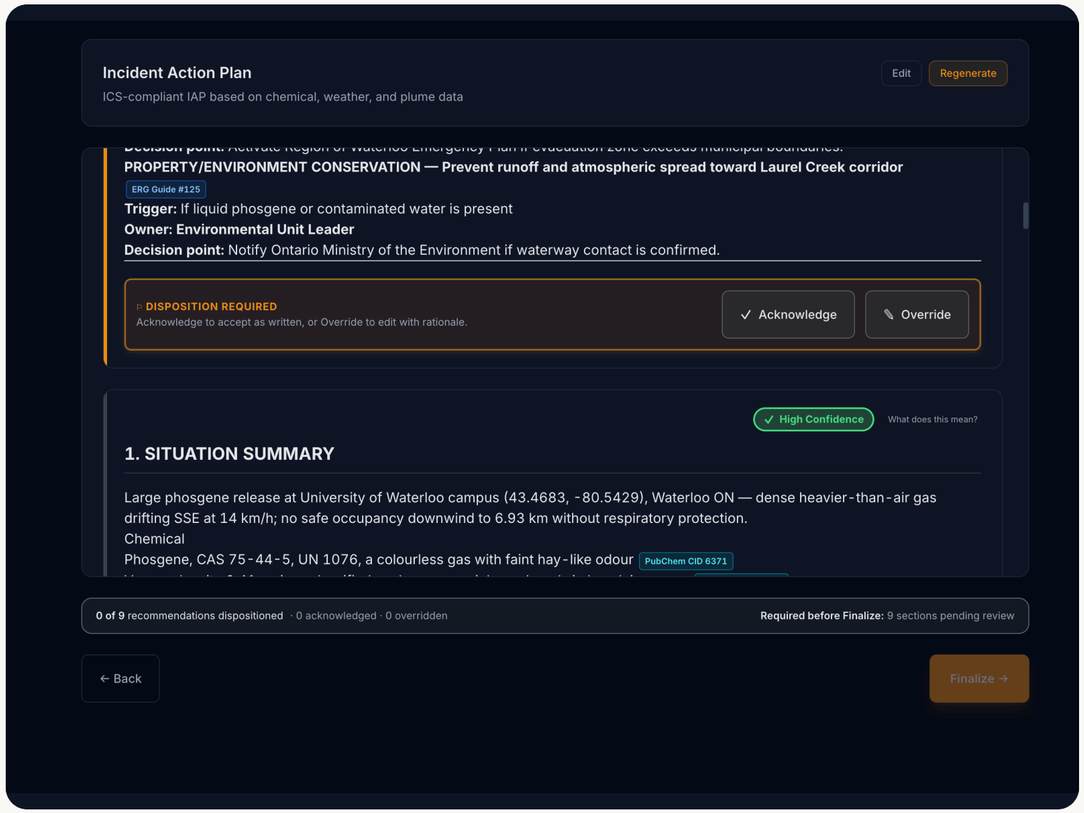
The tool says what it doesn’t know. That is a degraded mode, by design.

People do the sensemaking. The AI’s job is to say what changed.

Real screen, hazmat incident, test data. Same principle for a structure fire picture. · Weick 1993 · Klein et al. 2010 · Guyeux 2026 · both reviews

5 · FROM RESEARCH TO TOOLS

It shows its reasoning, and how sure it is



Trigger · Owner · Decision point

Every recommendation says why, who, and when to act. With its source.

“High confidence. What does this mean?”

An honest “how sure,” and a link that explains it.

Acknowledge, or override with a reason

0 of 9 dispositioned. You cannot finalize until a person has touched each one.

In a computer test, AI claims that passed an evidence check were right 71% of the time; raw AI output, 52%. No people tested yet.

Real screen, test data. · Berndt and Herczeg 2019 · Tan et al. 2026 · Goddard et al. 2012 · both reviews

From a chest strap to the commander’s screen

SENSE

Worn: heart rate, beat-to-beat variability, breathing rate, oxygen saturation, skin temperature.

Carried: fine particulate, volatile organics, exposure dose, position.

ESTIMATE

Fatigue, stress and a risk level, computed live for each firefighter.

Core temperature from a wearable is an estimate from heart rate, not a measurement.

SHOW

One board: crews sorted by distress first, then by risk. Vitals history one tap away.

Each firefighter can see their own numbers.

DECIDE

Rotate the crew. Send them to rehab. Pull the team.

The officer decides. The record keeps the reading and the decision.

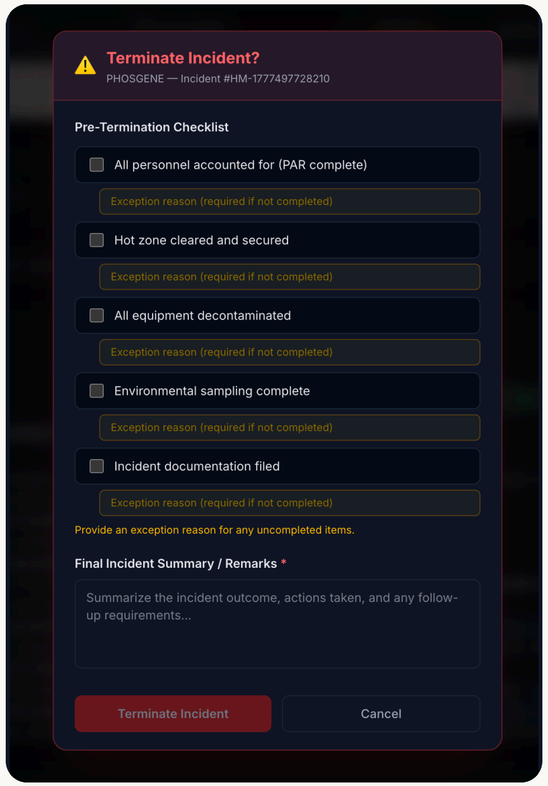
A model output. Validate it against ground truth before anyone acts on it.

One threshold the crews helped set, not ten alerts. Alerts become noise.

Never punitive, de-identified, voluntary, visible to the wearer. Or it stays in the locker.

Research prototype in development; no field evidence yet. · Buller et al. 2013 · Topazian et al. 2025 · Bai et al. 2025 · Ramey et al., systematic review

Real time interaction: The system asks. A person answers. The exception is recorded.



“PAR complete” is a box a person ticks
Not a status the system infers from a tracking board.

No exception without a reason
If a line is not complete, the reason is required, and it is kept.

Terminate stays locked until every line is answered
The check is in the task, not a signature at the end of the shift.

Same rule for a live accountability board: the system tracks, the officer accounts.

Real screen, test data. · Green 2022 · Goddard et al. 2012 · both reviews

What makes oversight real: Human in the loop

The person can judge

They have the information, the time and the training to see when the tool is wrong.

The person can disagree

They have the authority to override it, and sometimes they do.

Someone measures it

How much people rely on the tool, and how often they override it, is counted and reviewed.

Take any one away and “a human makes the final call” is a sentence in a policy, not a safeguard.

Synthesis of Green 2022, Goddard et al. 2012, Parasuraman and Manzey 2010, and both reviews

6 · EVALUATE BEFORE YOU ADOPT

Four questions, four instruments

How hard was that?

NASA-TLX. Six scales, 0 to 100: mental, physical and time demand, performance, effort, frustration.

“How mentally demanding was the task?”

2 minutes, after each run

Did they see what mattered?

SAGAT. Freeze the scenario, blank the screens, ask questions with known answers. Score against the truth.

“Which way is the wind now? How many crews are inside the line?”

Inside the run, two or three freezes

Would they use it again?

System Usability Scale. Ten statements, agree to disagree, scored 0 to 100. Average is 68.

“I think that I would like to use this system frequently.”

2 minutes, at the end

How much did they lean on it?

S-TIAS, three statements on trust in the AI, plus the log: every acknowledge and every override, with its reason.

The log is what replaces a feeling with a count.

1 minute, after each run

Hart and Staveland 1988 · Endsley 2021 · Lewis 2018 · McGrath, Lack, Tisch and Duenser 2025

6 · EVALUATE BEFORE YOU ADOPT · THE TWO-HOUR PILOT

One afternoon, one engine company, one classroom

Same scenarios with and without the AI · one run where it is wrong · everything logged

0:00
Brief
10 min
The scenarios, the tool, what you will measure, and that the tool may be wrong.

0:10
Run 1: without the AI
30 min
SAGAT freezes inside the run. NASA-TLX after. Decisions and times logged.

0:40
Run 2: with the AI
30 min
Same measures, plus S-TIAS. Every acknowledge and override, with its reason.

1:10
Run 3: the AI is wrong
30 min
Same measures. Did anyone catch it? How long did it take? What did they do?

1:40
Debrief and write-up
20 min
SUS, then four lines: who took part, how mature the AI is, the measures, the context.

If you never see it fail in training, you'll first see it fail on a real call.

Reporting items from Ramey et al., systematic review (in preparation)

6 · EVALUATE BEFORE YOU ADOPT

What the numbers tell you

If you see	It means	Do
Workload up with the AI	It added work: entry, alerts, a screen to feed.	Ask what it can leave off. Don’t buy until it goes down.
Awareness down with the AI	The screen buried what mattered.	Fewer layers, “what changed,” confidence up top.
Nobody caught the wrong run	Over-trust. Automation bias at work.	A check in the task, a named owner, train with it wrong.
Trust collapsed after the wrong run	Under-trust. One failure switched it off.	Show how sure it is; let people see its record.
Zero overrides	Nobody is checking.	Oversight is a sentence in a policy, not a safeguard.

Lewis 2018 · Parasuraman and Manzey 2010 · Dietvorst et al. 2015 · Green 2022 · Ramey et al., systematic review

6 · EVALUATE BEFORE YOU ADOPT

Six questions for any vendor

1. Tested with whom? Firefighters, in realistic conditions?
2. What happened to workload, awareness and trust? Measured how?
3. What does it do when it’s unsure, and how does it tell us?
4. How does a firefighter override it, and is the override recorded?
5. What data did it learn from, and whose workflow was it built around?
6. **Who is accountable when it’s wrong?**

A vendor who answers these well is worth a second meeting. A demo is not an answer to any of them.

Built from the gaps in Ramey et al., systematic review, and Ramey and Rehman, scoping review (draft)

6 · EVALUATE BEFORE YOU ADOPT

Every answer needs a chain of custody

Data it saw
Which pre-plan, which feed, timestamped.

→

Prompt
Yours, and the hidden one the vendor wrote.

→

Model and version
Swapped and updated without telling you.

→

Output
As produced, before anyone edited it.

→

Decision
Who accepted or overrode, and why.

Written once. Never edited. Yours to read, not the vendor’s to keep.

THREE CHECKS YOU CAN RUN IN ANY DEMO

1 **Ask it twice. Same answer?**

2 **Pick one claim. Trace it to a source that exists.**

3 **Switch a feed off. Does the screen tell you?**

Guyeux 2026 · Ramey and Rehman, scoping review (draft) · Interview study on governing AI in Canadian fire services (planned)

7 · TRY IT · 5 MINUTES

One tool, two questions, the person beside you

1. Pick one tool your department uses or is being pitched, AI or not. Or: a wearable that estimates fatigue · a drone that flags victims · a board that calls PAR by itself · AI that drafts the report.

2. **What is its wrong output, and who would catch it, and how?**

3. **Who is accountable when it is wrong?** A named role, never “the system.”

Evaluate it on your phone



thehumanfactor.ca/check/s2s

3 min in pairs | **2 to hear two**

CLOSE

Three things to do Monday

1 **Inventory**
What AI is already in use in your department, officially or not?

2 **Pilot before you buy**
Run the two-hour pilot, with the AI wrong at least once.

3 **Write the rule before go-live**
Who may override it, how that is recorded, who is accountable when it’s wrong, and what data about your people it keeps.

Human judgment is necessary, and not automatic.

AI will be in your department and at your incidents. The only question is whether your people shape it, or it shapes them.

If your department wants to be a field site for this research, I'd like to hear from you.

Scott Ramey

PhD Candidate, Systems Design Engineering, University of Waterloo

Slides, the safety checklist and every source

thehumanfactor.ca/s2s2026-practice

Sources

1 · What AI is

Moffatt v. Air Canada, 2024 BCCRT 149 · McCarthy et al. 1955 · Weizenbaum 1966 · Vaswani et al. 2017 · OpenAI 2022; UBS estimate via Reuters 2023 · Treasury Board of Canada Secretariat, Guide on the use of generative AI · Sharma et al. 2024, ICLR · OpenAI 2025 · Mata v. Avianca (S.D.N.Y. 2023) · Zhang v. Chen, 2024 BCSC 285 · Bloomberg, May 2023 (Samsung) · Microsoft and LinkedIn, Work Trend Index 2024

2 · People and AI

Bainbridge 1983 · Endsley 2023 · FCC 2018, Hawaii false alert report · Parush et al. 2020 · Bai et al. 2025 · Endsley 1995 · Hancock et al. 2011 · Dietvorst, Simmons and Massey 2015 · Zou et al. 2026 · Alter and Oppenheimer 2009 · Sundar 2008 · Nickerson 1998 · Tversky and Kahneman 1974 · Easterbrook 1959 · Klein 1989, 1998 · Klein, Calderwood and Clinton-Cirocco 1986 · Keinan 1987 · Staal 2004 · Parasuraman and Manzey 2010 · Goddard, Roudsari and Wyatt 2012 · Green 2022 · Lee et al. 2025 · Spearing et al. 2026 · Tan et al. 2026 · Pinero de Plaza et al. 2025

3 · What the research shows

Ramey, Masoumi-Ravandi, Rehman and Cao, systematic review (in preparation) · Ramey and Rehman, scoping review (draft) · Perry et al. 2012 · Topazian et al. 2025, JOEM · Topazian et al. 2026, New Solutions · Adams 2025, J Criminal Justice · CPSE Center for Innovation AI Strategic Scan 2025, via FireRescue1

4 · Safer, or not · 5 · From research to tools

Endsley 1995, 2023 · Parush et al. 2020 · Bai et al. 2025 · Hancock et al. 2011 · Dietvorst et al. 2015 · Zou et al. 2026 · Parasuraman and Manzey 2010 · Goddard et al. 2012 · Topazian et al. 2025 · Buller et al. 2013 · Weick 1993 · Klein, Wiggins and Dominguez 2010 · Guyeux 2026 · Berndt and Herczeg 2019 · Green 2022

6 · Evaluate before you adopt

Hart and Staveland 1988 (NASA-TLX) · Endsley 2021 (SAGAT and SPAM) · Lewis 2018 (SUS) · McGrath, Lack, Tisch and Duenser 2025 (S-TIAS) · Guyeux 2026 · Ramey and Rehman, scoping review (draft)

Full citations with DOIs: thehumanfactor.ca/s2s2026-practice