AI Capability Practice Guide: High-Risk or Public-Impact Contexts

Designing AI Use Where Mistakes Have Serious Consequences

Stage 1 — Orientation & Fast Start (High-Risk / Public-Impact)

Who This Guide Is For

This guide is for anyone working in **contexts where AI use can cause serious, wide, or long-lasting impact**, including:

- healthcare and clinical decision support
- public health communication and surveillance
- humanitarian response and crisis management
- public policy, regulation, and justice
- high-stakes education and assessment
- infrastructure, safety, and critical services
- research with significant public or ethical implications

You may be:

- a practitioner making frontline decisions
- a manager approving or shaping workflows
- a policy or ethics lead setting boundaries

- a governance or oversight body member

You **do not** need deep technical AI expertise.
You **do** need to make decisions that protect people, trust, and legitimacy.

Who This Guide Is Not For

This guide is **not** designed for:

- low-stakes personal productivity use
- informal experimentation with no real-world consequences
- tool-optimisation or “prompt hacking” for speed alone
- situations where errors are easily reversible

If mistakes in your context are easily fixed and affect only you, this guide is **overkill**.

If mistakes could:

- harm people
- mislead the public
- entrench inequity
- damage trust or reputation
- create legal or ethical exposure

…this guide is for you.

What You Will Be Able to Do in 30–60 Minutes

By working through this guide, you will be able to:

- recognise when an AI-related task has **high-risk or public-impact characteristics**
- decide whether AI is appropriate **at all** in that context
- design **safer human–AI workflows** for high-stakes tasks
- apply **ethics, equity, and harm-avoidance** before deployment
- set **clear escalation and "stop-use" conditions**
- document decisions in a way that can withstand scrutiny

You will also produce at least one **concrete safeguard artefact** (e.g. red-line rules, escalation triggers, or a high-risk AI use note) you can apply immediately.

FAST START — USE THIS NOW

If you are in a high-risk or public-impact context **today**, read this section first.

This Fast Start is designed to help you **avoid preventable harm** in **10 minutes**.

When to Use This Guide

Use this guide when:

- AI is being considered for:
    - clinical or care decisions
    - advice affecting safety, liberty, or rights
    - public-facing communication in crises
    - humanitarian triage or resource allocation
    - grading, selection, or exclusion decisions
- people may rely on outputs they cannot easily verify
- those affected have **limited power to contest or appeal**
- you would be uncomfortable if use became public knowledge
- you are unsure if AI should be used at all

If the answer to *"Could someone be seriously harmed or unjustly disadvantaged here?"* is **not clearly no**, treat the context as high-risk.

The 10-Minute High-Risk Entry Workflow

Use this sequence **before** you introduce AI into a high-risk / public-impact task **or** act on AI outputs in such tasks.

Step 1 — Name the High-Risk Decision or Impact

Write:

> *"This task could harm or significantly affect people because: [reason]."*

Examples:

- people may change treatment or behaviour based on this
- eligibility or access may be denied
- a community may be misinformed or stigmatised
- a crisis response may be redirected

If you cannot articulate the potential impact, **do not use AI yet**.

Step 2 — Decide if AI Is Appropriate At All

Ask:

1. **Is this primarily a judgement task?** (ethics, values, context)
2. **Would we be comfortable defending AI use in this context publicly?**
3. **Are there safer non-AI alternatives?**

If you cannot justify AI use **better than** non-AI alternatives, the default in high-risk contexts is:

> **Do not use AI, or restrict use to low-impact, clearly separated support (e.g. internal drafting only).**

Step 3 — Apply the High-Risk Co-Agency Rule

If you still consider using AI, define:

- **AI may only:** suggest wording, propose options, summarise background
- **AI may never:** make, frame as “correct”, or directly trigger decisions
- **Humans must:** verify, contextualise, and retain full responsibility

If this cannot be guaranteed in practice, **do not deploy AI**.

Step 4 — Run the Harm & Equity Screen

Ask:

1. **Who could be harmed if this output is wrong or incomplete?**
2. **Who is least able to challenge or correct it?**
3. **Could this deepen existing inequities or stigma?**
4. **Would harm likely be visible early—or only after damage is done?**

If you cannot mitigate identified harms **before use**, AI is **not yet safe to deploy**.

Step 5 — Set Escalation and Stop-Use Conditions

Before you proceed, define:

- **Escalation triggers:**
    - any serious complaint or incident
    - signs of bias or systemic harm
    - conflicting expert or community feedback

- **Stop-use triggers:**
    - evidence of significant harm or unfairness
    - systematic misinformation
    - loss of trust from key stakeholders

If you cannot commit to stopping or revising use under these conditions, governance is inadequate.

Worked Fast-Start Example — One Context, Three Paths

Context
Drafting public guidance on a health-risk topic using AI.

✅ Cautious, Acceptable Use (Support Only)

- AI helps generate **internal** drafts
- Experts verify facts, nuance, and implications
- Final wording is human-authored and owned
- AI use is clearly **supporting, not deciding**

Why this works:
AI accelerates drafting; humans retain expertise and responsibility.

⚠ Marginal, Fragile Use

- AI drafts are lightly edited for public release
- Limited expert review
- No documented harm or equity screen

Why this is fragile:
No clear evidence of safety or fairness; trust can collapse quickly if errors surface.

⛔ Unacceptable Use

- AI outputs are pushed directly to public channels
- No domain expert review
- No record of how guidance was generated

Why this fails:
People could be harmed by unverified information; no one can explain or own the decision.

Your First High-Risk Safeguard Artefact (Create This Now)

Write a **High-Risk AI Use Note** (5–7 lines):

- **Context:**
- **Potential harms:**
- **Who is most exposed:**
- **AI role (if any):**
- **Non-AI alternatives considered:**
- **Escalation / stop-use conditions:**

If you cannot fill this out convincingly, the safest move is **not to introduce AI** at this time.

How This Guide Works in High-Risk / Public-Impact Contexts

This guide prioritises:

- **safety over speed**
- **judgement over automation**
- **equity over efficiency**
- **accountability over convenience**

You are not expected to ban AI outright.
You are expected to:

- allow safe, tightly-bounded uses where justified
- block or redesign uses that create unjustifiable risk

- document and own decisions about AI introduction or removal

---

The Six Domains in High-Risk Work

In high-risk or public-impact domains, the six domains function as **protective layers**:

- **AI Awareness & Orientation** — Understanding failure modes and limits
- **Human–AI Co-Agency** — Ensuring humans remain fully responsible
- **Applied Practice & Innovation** — Allowing careful, reversible experimentation
- **Ethics, Equity & Impact** — Prioritising do-no-harm and justice
- **Decision-Making & Governance** — Making use explainable, reviewable, stoppable
- **Reflection, Learning & Renewal** — Improving safeguards after each cycle

Skipping domains in high-risk settings is not a time-saving tactic.
It is a **risk transfer** to those with least protection.

Stage 2 — How This Guide Works & Situational Entry Points

AI Capability Practice Guide: High-Risk or Public-Impact Contexts

---

How This Guide Is Designed to Be Used

This guide is **not** meant to be read linearly or absorbed as theory.

It is designed for:

- moments of uncertainty
- high-stakes decisions
- partial information
- institutional pressure
- ethical ambiguity

You are expected to:

- enter where the situation demands
- apply only the domains required
- document what you decide
- move on

The goal is **risk containment and legitimacy**, not technical mastery.

Why High-Risk Contexts Require a Different Usage Pattern

In high-risk or public-impact contexts:

- speed increases harm, not productivity
- confidence can be misleading
- reversibility is limited
- accountability cannot be delegated
- trust is fragile and cumulative

Therefore:

- **AI capability here is primarily defensive**
- innovation must remain **careful, bounded, and reviewable**
- refusal to use AI is sometimes the **most responsible decision**

This guide helps you decide:

when not to use AI
how to restrict it safely when you do
how to justify those decisions later

The Guide at a Glance

The guide is structured around **three layers**:

Layer 1 — Immediate Safeguards

- Fast Start workflow
- Harm & equity screening
- Stop-use and escalation logic

Layer 2 — Capability Workflow

- Six domains, applied as protective checks
- Focus on judgement, not automation

Layer 3 — Institutional Defensibility

- Documentation cues
- Review and audit readiness
- Public justification mindset

You will move between layers depending on urgency and risk.

The Six Domains as Protective Controls

In high-risk environments, the six domains act less like a growth pathway and more like **safety barriers**.

Domain	Primary Function in High-Risk Contexts
Awareness & Orientation	Identify where AI is likely to mislead or hallucinate
Human–AI Co-Agency	Prevent implicit delegation of judgement
Applied Practice & Innovation	Allow limited, reversible support only
Ethics, Equity & Impact	Anticipate harm before it manifests
Decision-Making & Governance	Ensure explainability, challengeability, and stop-use
Reflection & Renewal	Adapt safeguards after incidents or near-misses

Skipping one barrier does not save time.
It simply shifts risk to people with fewer protections.

Situational Entry Points

Do **not** start with a domain.
Start with **the situation you are in**.

Each entry point below tells you:

- which domains matter first
- what action to take immediately
- what failure looks like

Entry Point 1 — "Someone could be harmed."

You believe:

- safety, wellbeing, rights, or dignity are at stake
- outcomes may affect vulnerable individuals or groups

Apply first:

- Ethics, Equity & Impact
- Decision-Making & Governance

What to do now

- Identify affected groups explicitly
- Assume limited ability to contest decisions
- Increase review thresholds
- Define stop-use conditions

Common failure mode

- “No one has complained yet” used as justification

Entry Point 2 — “This will be public.”

The output will be:

- published
- broadcast
- distributed at scale
- relied upon by non-experts

Apply first:

- AI Awareness & Orientation
- Human–AI Co-Agency

What to do now

- Identify what the system cannot reliably know
- Restrict AI to background or draft-only roles
- Ensure final ownership is human and declared

Common failure mode

- Equating polish with reliability

Entry Point 3 — “The decision seems small, but consequences aren’t.”

The task feels routine, but:

- consequences compound
- precedents may form
- downstream effects are unclear

Apply first:

- Awareness & Orientation
- Ethics, Equity & Impact

What to do now

- Map downstream effects
- Identify where harm might appear late
- Treat early signals seriously

Common failure mode

- Normalising incremental risk

Entry Point 4 — “Experts disagree.”

You observe:

- conflicting professional judgement
- unclear evidence
- contested values

Apply first:

- Human–AI Co-Agency
- Decision-Making & Governance

What to do now

- Prevent AI from “averaging away” disagreement
- Use AI only to surface alternatives, not resolve them
- Preserve human deliberation

Common failure mode

- Using AI to artificially manufacture certainty

Entry Point 5 — “We’re under pressure to move fast.”

Pressure comes from:

- crisis response
- public scrutiny
- political or reputational urgency

Apply first:

- Co-Agency
- Ethics, Equity & Impact

What to do now

- Slow the decision moment, not the work
- Narrow AI's role
- Increase human oversight

Common failure mode

- Speed used to justify reduced scrutiny

Entry Point 6 — "This might be reviewed later."

You expect:

- audit
- inquiry
- legal challenge
- media scrutiny

Apply first:

- Decision-Making & Governance
- Reflection & Renewal

What to do now

- Document decision logic now
- Make assumptions visible
- Preserve evidence of review and restraint

Common failure mode

- Reconstructing decision logic after the fact

How to Use This Guide Under Severe Time Constraints

If you have 5–10 minutes

- Use the Fast Start
- Run the harm & equity screen
- Decide whether to pause, restrict, or proceed
- Write a minimal use note

This alone prevents many failures.

If you have 20–30 minutes

- Identify the most relevant entry point
- Apply 2–3 domains deliberately
- Define escalation and stop-use triggers

This is the minimum for serious public-impact work.

If stakes are very high

- Apply all six domains
- Require domain-specific expertise
- Treat restraint as a valid outcome

- Prioritise legitimacy over output speed

What Comes Next

From here, the guide moves into the **core practice workflow**, applying the six domains **as safeguards**, not innovation accelerators.

Stage 3 focuses on:

- **Domains 1–3**
- understanding AI limits
- designing strict co-agency
- permitting only safe, reversible AI support

Stage 3 — Core Practice Workflow: Domains 1–3

AI Capability Practice Guide: High-Risk or Public-Impact Contexts

Domains 1–3 determine **whether AI should be present at all**, and if so, **how tightly its role must be constrained**.

In high-risk or public-impact contexts:

- these domains are not about creativity
- they are about *preventing silent failure*
- they establish **non-negotiable boundaries**

Domain 1 — AI Awareness & Orientation

Knowing When AI Is Likely to Mislead

What This Domain Protects

This domain protects against:

- over-confidence in fluent outputs
- hallucinated facts or false certainty
- hidden assumptions presented as neutral truth
- misuse of AI beyond its evidentiary limits

In high-risk contexts, misunderstanding AI behaviour is a **direct harm vector**.

Apply Now — Critical Awareness Questions

Before AI use (or reliance on output), ask:

- What type of system is this (generative, predictive, decision-support)?
- What information is *likely missing* from its training or context?
- Where could it confidently be wrong?
- What error would cause the most harm?

If you cannot articulate **specific failure modes**, AI is being over-trusted.

Tool in Use — High-Risk Awareness Check

Use this short diagnostic:

High-Risk Awareness Check

- This system is good at:
- This system is *not* reliable for:
- A dangerous-but-plausible error here would be:
- Human expertise required at these points:

One sentence per line is enough.

Common Failure Modes

- Treating AI summaries as factual synthesis
- Assuming neutrality in morally loaded contexts
- Ignoring data gaps that affect marginalised groups
- Allowing “confidence tone” to stand in for evidence

Quick Reflection

> *What would a “reasonable-sounding but wrong” output look like here?*

If you cannot imagine this, you are not yet safe to use AI.

Domain 2 — Human–AI Co-Agency

Preventing Delegation of Judgement

What This Domain Protects

This domain protects:

- accountability
- ethical ownership
- professional duty of care

In high-risk contexts, **judgement must never migrate to AI by default**.

Apply Now — Co-Agency Boundary Questions

Ask explicitly:

- What specific support is AI allowed to provide?
- Which decisions are **non-delegable**?
- Who is accountable if harm occurs?

If responsibility becomes diffuse, co-agency has failed.

Tool in Use — Restricted Co-Agency Map

Define roles in writing:

Restricted Co-Agency Map

- AI may: draft, list alternatives, summarise background
- AI may *not*: decide, prioritise, diagnose, approve, or conclude
- Human role: interpret, verify, decide, and own consequences

If exceptions exist, document them.

Common Failure Modes

- “Rubber-stamping” AI outputs under time pressure
- Allowing AI to frame options in biased ways
- Treating generated recommendations as defaults

Quick Reflection

> *If this decision were challenged, could I clearly separate what AI did from what humans decided?*

If not, accountability is already compromised.

Domain 3 — Applied Practice & Innovation

Allowing Only Safe, Reversible Support

What This Domain Enables (Carefully)

This domain allows:

- limited experimentation
- cautious AI support
- learning in controlled environments

In high-risk contexts, innovation must be:

- reversible
- reviewable
- contained

Apply Now — Safe Practice Questions

Before experimenting, ask:

- Can this use be reversed if wrong?
- Will anyone act on this output directly?
- Are consequences contained or cascading?

If outputs can trigger irreversible action, experimentation is inappropriate.

Tool in Use — Safe-Use Prompt Frame

When AI is used, explicitly constrain it:

> "Generate a draft or set of options for internal consideration only.
> This output will be reviewed by domain experts before any action or publication."

This framing reduces authority creep.

Common Failure Modes

- Treating pilot use as low-risk by default
- Scaling before understanding impact
- Confusing internal drafts with decision support

Quick Reflection

> *Am I testing AI capability—or testing public tolerance for failure?*

If it's the latter, stop.

Domains 1–3 Combined — The High-Risk Gate

Before progressing beyond internal use, all three must be satisfied:

- AI failure modes are understood
- Human responsibility is explicit
- Use is strictly bounded and reversible

If any are missing, **do not proceed**.

This is not conservatism.
It is **professional duty** in high-impact work.

Transition to Stage 4

Domains 1–3 decide *if and how* AI can appear.
Domains 4–6 determine *whether its presence is ethically, socially, and institutionally defensible*.

Stage 4 focuses on:

- harm prevention
- equity and justice
- governance, escalation, and renewal

Stage 4 — Risk, Responsibility & Renewal: Domains 4–6

AI Capability Practice Guide: High-Risk or Public-Impact Contexts

Domains 4–6 are where **most real-world harm is either prevented or amplified**.

In high-risk or public-impact settings, these domains are **not optional safeguards**.
They are the difference between:

- responsible restraint
- and institutional failure

Domain 4 — Ethics, Equity & Impact

Preventing Harm Before It Occurs

What This Domain Protects

This domain protects:

- people affected by AI-mediated decisions
- communities with limited power to contest outcomes
- public trust in institutions
- ethical legitimacy over time

In high-risk contexts, ethics is not about intent.
It is about **foreseeable harm and unequal exposure**.

AI can:

- scale bias rapidly
- conceal value judgements
- normalise inequitable trade-offs

Ethical capability means **anticipating harm before deployment**, not managing fallout afterwards.

Apply Now — Harm & Equity Questions

Before authorising AI use, ask:

- Who could be harmed if this output is wrong, incomplete, or misleading?
- Which groups are most exposed — and why?
- Are some people less able to challenge or correct outcomes?
- Would harm surface early, or only after damage is done?
- Is there moral disagreement hiding behind technical framing?

If harms cannot be mitigated *before* use, pause or prohibit AI.

Tool in Use — High-Risk Ethical Impact Scan

Use this for any public-facing or high-stakes use:

Ethical Impact Scan (High-Risk)

- Potential harms:
- Most affected groups:
- Bias or exclusion risk:
- Reversibility of harm:

One sentence per line is sufficient — clarity matters more than length.

Common Failure Modes

- assuming accuracy guarantees fairness
- prioritising efficiency over dignity
- overlooking downstream or indirect harm
- treating equity as an optional lens
- deferring ethics to legal review alone

Quick Reflection

> *If the worst plausible outcome occurred, who would pay the price — and would we accept that?*

Domain 5 — Decision-Making & Governance

Making AI Use Explainable, Contestable, and Stoppable

What This Domain Protects

This domain protects:

- accountability and traceability
- legal and regulatory legitimacy
- institutional credibility
- the right to challenge and appeal
- the ability to halt use when harm emerges

In high-risk contexts, **undocumented AI use is a governance failure**.

Apply Now — Governance Questions

Before approving or continuing AI use, ask:

- Who authorised this use, and under what conditions?
- How is AI influence documented?
- What review or escalation mechanisms exist?
- Who can stop use if harm is identified?
- How would this decision be explained publicly?

If you cannot answer these clearly, governance is inadequate.

Tool in Use — High-Risk Decision Record

Create a record *before* deployment:

High-Risk AI Decision Record

- Context and purpose:
- Role of AI:
- Human decision authority:
- Risks and mitigations:
- Escalation and stop-use triggers:

This record is not bureaucracy.
It is **permission to proceed responsibly**.

Escalation & Stop-Use Triggers

At minimum, establish:

Immediate Escalation When:

- credible harm or bias is reported
- experts raise unresolved concerns
- real-world effects diverge from expectations

Immediate Stop-Use When:

- people are harmed or misled
- misinformation spreads at scale
- trust collapses among affected groups

If stop-use feels unrealistic, the AI use is not safe.

Common Failure Modes

- treating governance as post-hoc justification
- failing to empower anyone to halt use
- fragmented oversight across teams
- normalising exception-making under pressure

Quick Reflection

> *Who can stop this use today — in practice, not theory?*

Domain 6 — Reflection, Learning & Renewal

Ensuring Safeguards Improve Over Time

What This Domain Sustains

This domain sustains:

- institutional learning
- ethical maturity
- adaptive safeguards
- credibility after incidents
- long-term trust

In high-risk environments, learning is often reactive.
Capability requires **deliberate reflection even when nothing "goes wrong."**

Apply Now — Reflection Questions

After AI use or decisions, ask:

- What risks emerged that we did not anticipate?
- Which safeguards worked — and which didn't?
- Were early warning signs missed or ignored?
- What needs to change before next use?

Without reflection, near-misses become future failures.

Tool in Use — Safeguard Renewal Loop

Use this simple cycle:

Observe → Adjust → Re-Authorise

- Observe outcomes and signals
- Adjust boundaries or controls
- Re-authorise use only if conditions still hold

This ensures use remains conditional, not permanent by default.

Common Failure Modes

- learning only after crises
- failing to update policies post-incident
- assuming initial approval is sufficient
- prioritising continuity over safety

Quick Reflection

> *What assumption about "acceptable risk" needs revisiting?*

Domains 4–6 Combined — The Public-Impact Test

Before ongoing or scaled AI use, confirm:

- harms have been meaningfully considered and mitigated
- governance is active, documented, and empowered
- learning mechanisms are in place

If any condition is missing, **do not scale**.

Responsible restraint is a form of capability.

Transition to Stage 5

Stage 5 consolidates this work into:

- a **High-Risk Capability Self-Check**
- a **worked public-impact scenario**
- a **personal or organisational operating model**

This final stage ensures readiness under pressure.

Stage 5 — Capability Self-Check, Worked High-Risk Scenario & Operating Model

Stage 5 turns this guide into a **fully operational system** you can use in any high-risk or public-impact environment.

You will:

- assess your current high-risk capability
- walk through a realistic scenario end-to-end
- build a **High-Risk AI Operating Model** you can apply immediately

This stage ensures **practical readiness**, not theoretical understanding.

PART A — HIGH-RISK CAPABILITY SELF-CHECK

This is not an audit and not a score.
It orients you to **where harm could arise if AI is misused or misunderstood**.

Complete in **5–7 minutes**.

Domain 1 — Awareness & Orientation

Consider:

- I can identify situations where AI is likely to be confidently wrong
- I know what information AI cannot reliably access or interpret
- I can explain limits and uncertainty in plain language

☐ Mostly yes ☐ Mixed ☐ Mostly no

If "mixed/no":
AI should *not* be used in any public-impact decisions until clarity improves.

Domain 2 — Human–AI Co-Agency

Ask:

- I define what AI may and may not do in high-risk work
- I ensure humans—not AI—retain all final decision authority
- Accountability is explicit and non-transferable

☐ Mostly yes ☐ Mixed ☐ Mostly no

If “mixed/no”:
You are exposed to responsibility drift.

Domain 3 — Applied Practice & Innovation

Reflect:

- Any AI experimentation I allow is reversible
- No one acts on AI output without human verification
- I avoid “pilot creep” into ungoverned deployment

☐ Mostly yes ☐ Mixed ☐ Mostly no

If “mixed/no”:
You are operating with unmanaged risk.

Domain 4 — Ethics, Equity & Impact

Evaluate:

- I can identify who is most exposed to harm
- I recognise when AI may deepen inequity
- I pause use when effects are unclear or non-reversible

☐ Mostly yes ☐ Mixed ☐ Mostly no

If “mixed/no”:
Ethical risk is high and unmitigated.

Domain 5 — Decision-Making & Governance

Check:

- AI influence is documented before decisions are made
- Escalation and stop-use triggers are defined
- Decisions would withstand public scrutiny

☐ Mostly yes ☐ Mixed ☐ Mostly no

If “mixed/no”:
Governance is not ready for high-risk AI.

Domain 6 — Reflection, Learning & Renewal

Ask:

- I review AI-influenced decisions for early warning signs
- Safeguards evolve as new risks emerge
- Lessons are shared, not siloed

☐ Mostly yes ☐ Mixed ☐ Mostly no

If "mixed/no":
Capability will degrade over time.

Interpreting Your Self-Check

- Gaps in Domains 1–2 → pause high-risk AI use immediately
- Gaps in Domains 4–5 → strengthen safeguards before authorising use
- Gaps in Domain 6 → reinvest in reflection and renewal

High-risk capability is not about skill.
It is about **containment, judgement, and restraint**.

PART B — WORKED HIGH-RISK SCENARIO (END-TO-END)

A realistic example showing the domains in action from start to finish.

Scenario

A government health agency considers using AI to **draft public guidance** about a rapidly spreading infectious threat.

The guidance will influence:

- public behaviour
- healthcare-seeking decisions
- anxiety and trust
- resource utilisation

This context is **high-impact, asymmetric, and potentially irreversible**.

Domain 1 — Awareness in Action

The team recognises:

- AI may hallucinate or oversimplify scientific uncertainty
- Training data may not reflect current epidemiology
- Polished language does not equal accuracy

Therefore:

- AI is restricted to **internal drafting**, not external publication
- Outputs are treated as **ideas to interrogate**, not facts to publish

---

Domain 2 — Co-Agency in Action

Roles are set clearly:

AI may:

- suggest structure
- generate alternative framings
- summarise background literature

AI may *not*:

- produce final public wording
- imply certainty
- advise on behavioural actions

Humans must:

- verify scientific accuracy
- contextualise with local data
- approve final outputs
- take full responsibility

Domain 3 — Applied Practice in Action

The team:

- tests multiple AI framings for clarity
- uses AI only as a **support tool**
- ensures all outputs remain reversible and non-binding

No content is released without expert intervention.

Domain 4 — Ethics & Impact in Action

Before drafting final guidance, the group examines:

- Who could be harmed by incorrect emphasis or omissions
- Which communities are most vulnerable to misinformation
- Whether language could stigmatise certain groups
- How misunderstanding might escalate risk behaviour

They alter content accordingly.

Domain 5 — Governance in Action

A **High-Risk AI Decision Record** is created:

- purpose of AI use
- scope and boundaries
- human authority
- risk mitigations
- stop-use triggers

Legal, ethical, and communication leads review and endorse.

Domain 6 — Reflection & Renewal in Action

After guidance is published, the agency:

- monitors complaints, misinformation spread, and behavioural outcomes
- identifies whether AI contributed to misinterpretation
- adjusts future guidance development
- revises safeguards and approval steps

Capability strengthens for future events.

What This Scenario Shows

Responsible high-risk AI use is defined by:

- restricted scope
- human accountability
- ethical foresight
- transparent documentation
- readiness to stop, revise, or withdraw

This is **capability**, not caution.

PART C — HIGH-RISK AI OPERATING MODEL

Use this to operationalise your high-risk practice.
Complete once, review periodically.

---

1 My High-Risk Contexts

Situations where AI requires extreme caution:

---

2 My AI Boundaries (Non-Negotiable)

AI may support:

AI must not:

---

3 Harm & Equity Red Lines

We pause AI use when:

We stop AI use immediately when:

---

4 Decision Documentation Requirements

For any AI-influenced public-impact decision, we will record:

- purpose and context
- AI role
- human authority
- risk considerations
- escalation triggers

---

5 Escalation & Stop-Use Protocol

Escalate when:

- concerns arise from experts or communities
- outputs diverge from evidence
- unintended effects appear

Stop use when:

- harm is observed
- trust collapses
- outputs mislead materially

---

6 Reflection & Renewal Rhythm

After major AI-influenced decisions, we ask:

- What worked?

- What caused concern?
- What will we change next time?

Revisit quarterly or after any incident.

THE HIGH-RISK COMMITMENT

I will not delegate judgement or responsibility to AI.
I will treat safety, dignity, and equity as non-negotiable.
I will design AI use to be reversible, reviewable, and stoppable.
I will learn from every use to improve future practice.