

# Commissioning Learning & Development in the UK Public Sector

*A 2026–27 guide for people who buy training.*

By EnlightenWorks ~15 min read Reviewed August 2026

## IN SHORT

- If you commission training and you are not a learning specialist, this is for you. It will not turn you into one; it will help you buy well.
- It covers four things: the context you are commissioning into, whether training is even the right answer, how to recognise good work in a proposal, and what to put in the specification.
- Everything here is a buyer's tool: checklists, questions and red flags, not a design manual.

---

## Who this is for

You have been handed a budget and a problem: a team that needs to work differently, a new duty to land, a capability gap someone has named. A supplier will design the training. Your job is to commission it well: scope it, specify it, choose between bids, and judge whether what came back is any good.

You do not need to know how to design a course to do that well. You need to know what good looks like, what to ask for, and what to be wary of. That is all this document is. It assumes you know your own governance, and assumes no learning-design knowledge.

---

## **Where this guide stops: procurement law**

This is a guide to judging the quality of what you buy, not to the law of buying it. Most procurements started in England, Wales and Northern Ireland since 24 February 2025 run under the Procurement Act 2023 and the Procurement Regulations 2024; Scotland has its own regime. Thresholds, frameworks and dynamic markets, transparency notices, conflicts, social value and your standing orders sit with your procurement team.

One rule from that world matters throughout this guide: award criteria and scoring must be clear, proportionate, related to the contract and disclosed in advance. Anything here you intend to score, including the ten markers and the seventeen questions, belongs in your published evaluation methodology, not in an undisclosed checklist.

---

# Before you commission anything: is training the answer?

The most expensive commissioning mistake is buying a course for a problem a course cannot fix. Done well, training changes what people know, what they can do, and how they understand and approach the work. What it cannot change is the conditions around the work. If people already know what to do but do not do it, the cause is usually somewhere else: the process is unclear, the tools are poor, the incentives point the other way, or no one has the time. A course will not touch any of that, and you will have spent the budget proving it.

Before you write a line of specification, ask:

1. **What, specifically, would people be doing differently** if this worked? Name a behaviour, not a topic.
2. **Do they not do it now because they cannot, or because something stops them?** If something stops them, that is not a training problem.
3. **If I changed only the conditions around the work and trained no one, would the problem shrink?** If yes, fix the conditions first, or alongside.

A good supplier will ask you these questions unprompted, and will sometimes tell you that you do not need them.

## **RULE OF THUMB**

Training is most likely to help where people lack knowledge, skill, practice or confidence, and have the support to apply it. Where the blocker is elsewhere (unclear process, wrong tools, weak incentives, no time), commission the fix, not a course.

---

# The 2026–27 context you are buying into

*Working assumptions as of August 2026. Check each against your own organisation; several are contested and move quickly.*

- **Hybrid is the default, and still argued over.** Most public sector organisations run some form of hybrid working; in the civil service, cross-government guidance sets a minimum expectation of 60% of working time in the office or on official business, while much of the operational workforce was never remote. *Implication:* reject a design that assumes everyone is co-located, and one that assumes everyone is remote. Ask how it earns its live time.
- **AI has arrived unevenly.** Increasingly approved on public sector systems, but policy varies by organisation and data classification. *Implication:* expect suppliers to use AI, and require them to say how, and to show they verify what it produces.
- **Delivery is place-based.** Public sector workforces are spread across sites, shifts and regions; in central government, Places for Growth had relocated more than 18,000 roles out of London by March 2024, ahead of its 22,000-by-2027 target. *Implication:* cohorts are often distributed; a design for one co-located group may not survive your workforce.
- **Every career stage, one workforce.** The same cohort can hold a new starter and a colleague with thirty years of service. *Implication:* tone, examples and formats must work across all of them, not only the most confident or the most online.
- **Budgets are tight and outcomes-led.** *Implication:* require an evaluation plan in the bid. If a supplier cannot tell you how you will know it worked, you cannot defend the spend.
- **Integrity is in the spotlight.** The first volume of the Post Office Horizon IT Inquiry's final report (July 2025), the Covid-19 Inquiry's rolling module reports (the July 2026 one covers pandemic procurement) and the Ministry of Defence's Afghan data breach, made public in 2025, have raised the stakes of ethics, decision-making and data-handling content. *Implication:* examples handled with care; impartiality is not optional.
- **Neurodiversity is a design assumption.** A commonly used estimate is that one in seven people are neurodivergent; the figure varies with the definitions used.

*Implication:* accessibility is a requirement you specify up front, not an adjustment bolted on later.

- **The training market is shifting.** In January 2026 the government announced a National School of Government and Public Services, and most parts of the public sector have a national offer of their own: the Government Campus and Civil Service Curriculum in central government, the College of Policing in policing, the NHS learning platforms in health. *Implication:* check what already exists in your sector before commissioning something new. You may be buying a duplicate.

---

# What good looks like, so you can recognise it in a bid

You do not need to design the training to judge a proposal. These are the ten markers of well-designed work; a strong bid shows most without being asked.

1. **It starts with behaviour, not topics.** Objectives say what a learner will be able to *do*, with a condition and a standard. "Understand" is too vague on its own; look for an observable demonstration (explain, apply, compare, diagnose). Its overuse is a tell.
2. **It is designed backwards from the outcome.** Decide the evidence first, design the experience to produce it. Slides come last.
3. **It shows a nuanced appreciation of the learner journey.** The design works within your time constraints and organisational context, and thinks about what happens before, what follows after, and what the learner keeps using.
4. **Every live minute is justified.** Synchronous time is spent on what needs people together: practice with feedback, real-time discussion, emotionally weighty content.
5. **It produces something people keep.** A job aid or template that outlives the session.
6. **It is accessible by design.** WCAG 2.2 AA, captions, plain English, inclusive examples, built in from the start.
7. **It is evaluated at the level of behaviour.** How you will know, at 30–90 days, whether anything changed.
8. **It is honest about what it cannot do.** A proposal with no risks listed has not been thought through.
9. **It is specific to your context.** Generic content that could be sold to anyone probably was.
10. **A named human stands behind it.** Someone puts their name to the design and the facts, alongside the organisation's own quality assurance.

**RULE OF THUMB**

Score a proposal 0–2 on each marker, as a working heuristic rather than a validated threshold. Below about 12 out of 20, ask the supplier to strengthen the weak areas first; if the score will affect the award, put it in your published evaluation methodology. It is cheaper to fix a bid than a delivered course.



# Telling genuine expertise from AI slop

Many suppliers now use AI somewhere in production, and that is fine: used well, it makes good designers faster and sharper. The problem is not AI. It is **AI slop**: generic, unchecked, outcome-free content dressed up as expertise. As a commissioner you need to tell one from the other, because you are paying for judgement, not for text.

The tell is not whether AI was used. It is whether **human judgement, verification and context** are visible in the work.

*Green flags versus slop, at a glance.*

| Expertise (AI-assisted or not)                                       | Slop                                                         |
|----------------------------------------------------------------------|--------------------------------------------------------------|
| ✓ Objectives with verbs, conditions, standards, tied to your problem | ✗ Objectives full of "understand", "be aware of"             |
| ✓ Specific, current facts, visibly checked and sourced               | ✗ Confident statistics with no source                        |
| ✓ Named trade-offs and honest limitations                            | ✗ No risks or limits listed                                  |
| ✓ An evaluation plan that could embarrass them if it failed          | ✗ A happy-sheet, or no evaluation at all                     |
| ✓ Examples that could only be about your context                     | ✗ Content that fits anyone, names swapped                    |
| ✓ A named person accountable                                         | ✗ Design myths as fact: learning styles, the 10%/90% pyramid |

This is measurable. In July 2026 Dr Philippa Hardman tested four frontier AI models against a 36-point instructional design standard: execution quality varied widely, design judgement scored far lower in every model, and not one questioned whether a course was the right solution. Three of the four also introduced errors of omission (required duties and escalation routes silently dropped) that a fact-accuracy review

would not catch ([Hardman, \*AI Can Build Your Course, but Can it Design the Learning?\*, July 2026](#)).

## Signs a supplier is at the leading edge

The flags above are the baseline. A smaller group of suppliers is further ahead in how they work with AI, and it shows when they talk about their own practice. None of these is essential, but each is a good sign that you are dealing with an informed, adaptive user rather than a subscriber:

- **They compare models, and can say why.** They talk about specific models and labs, and why they use one for drafting and another for checking: reasoning strength, context length, cost, data terms. "We use ChatGPT" is not an answer at this level.
- **They know when open weights matter.** They can explain the difference between hosted frontier models and open-weight models run under their own control, and when data sensitivity makes the second the right call.
- **They think in harnesses, not prompts.** They describe the workflow around the model: staged pipelines, review gates, structured checks, human sign-off points. Not one person typing into a chat window.
- **They test rather than eyeball.** They keep evaluation sets or check protocols for AI-assisted work, and can show you the failures they have caught.
- **They are close to the field.** They know what has changed in the last few months, which updates affected their pipeline, and what they adapted as a result. Model releases land monthly; a supplier whose AI practice has not changed in a year is not paying attention.
- **They know where it fails.** They can name tasks they have stopped using AI for, and what it still gets wrong in their domain.
- **They use AI beyond production.** For practice and feedback (rehearsal, role-play, spaced follow-through), not only for making content faster.
- **Their data answers are immediate.** Which tier of service, what agreements are in place, where data goes, what is never entered: answered without needing to check.

**RULE OF THUMB**

Ask one question: *"What in this proposal did a human decide, verify, or take responsibility for?"* If the honest answer is "very little", you are being sold slop at a specialist's price.

The paradox worth naming to your panel: good AI-assisted work and pure slop can look similar at a glance. The difference shows up in the specifics, the verification and the accountability. Commissioning well means looking there, not at the polish.

---

# What to put in your specification

A tight spec gets you better bids and a fairer evaluation. Each line below can be lifted into a tender.

## Purpose and outcomes

- The behaviour change you are buying, stated as an outcome, not a topic.
- Confirmation that training is the right intervention, or the parts of the problem it will and will not address.

## Design and delivery

- Objectives written with a verb, a condition and a standard; "understand" only with an observable demonstration attached.
- A design that reflects the learner journey, your time constraints and the organisational context, with follow-through and a lasting job aid.
- A justification for every hour of synchronous time.

## Accessibility, inclusion and the law

- WCAG 2.2 AA specified as your minimum standard for digital content.
- Where learning is delivered through a public sector website, intranet, extranet or in-scope mobile app, the Public Sector Bodies Accessibility Regulations 2018 may apply, normally requiring WCAG 2.2 AA and an accessibility statement. Reasonable-adjustment duties arise under the Equality Act 2010 in Great Britain and the Disability Discrimination Act 1995 in Northern Ireland. Confirm the position for your platform and organisation.
- Plain English; acronyms defined; inclusive, representative examples.
- Impartiality: content complies with your organisation's code of conduct and presents contested issues accurately, fairly and without party-political advocacy.
- UK GDPR and the Data Protection Act 2018 handling for learner data, with a retention policy.

- The commissioning body's applicable Welsh Language Standards and compliance notice, and any contractual language requirements.

## **AI use by the supplier**

Many suppliers now use AI somewhere in production. The spec should not ban it; it should set the terms, and it should keep legal duties separate from the controls you choose to impose.

- A statement of how AI is used in producing your materials, and how AI-generated content is verified against live sources before it reaches you.
- A review of AI-assisted content that checks for omission as well as accuracy: what the source material requires that the module has silently dropped.
- A tool inventory on request (under confidentiality where needed): which AI tools and models sit in the production pipeline, and what they process.
- Data protection law: where the supplier processes personal data on your behalf, UK GDPR (as amended by the Data (Use and Access) Act 2025) requires a written data processing agreement, and any international transfer needs a lawful route such as adequacy or safeguards.
- Buyer-side controls, set as contract terms rather than legal duties: named sub-processors, stated data residency, and no use of your data to train models without written agreement.
- Government information handled in line with the Government Security Classifications and your organisation's approved systems and handling rules: classified material only in tools approved for it.
- A named person accountable for the accuracy of AI-assisted content, alongside the supplier's own quality assurance, and a remedy where generated errors surface after delivery.
- Productivity assumptions: bidders state where they expect AI to create a material saving and how that affects price, timetable and quality. Judge the whole offer on value; a discount for AI use is not the point.

## **Evaluation, commercials and rights**

- An evaluation plan in the bid: reaction, learning and behaviour as a minimum, with an attempt at results and two leading indicators.

- Intellectual property and licensing: who owns the materials, and your right to reuse them.
- Named accountability for the design and its factual accuracy; quality assurance before delivery.

---

## Seventeen questions to ask a supplier

Sharp questions expose weak proposals faster than a scoring grid. If the answers will affect the award, build the questions and their scoring into your published evaluation methodology rather than applying them as undisclosed criteria.

1. What must a learner be able to *do* differently after this, and how will we see it?
2. Is training the right fix here? What part of the problem will it not touch?
3. How does the design fit our learner journey, our time constraints and the context it will land in: what happens before and after the day itself?
4. Which parts of the live time could not be done any other way, and why?
5. What will a learner still be using in three months?
6. How have you designed this for someone neurodivergent, using a screen reader, or joining from home?
7. How will we know, at 60 to 90 days, whether behaviour changed?
8. Which of your success measures could look good even if the course failed?
9. How did you use AI in producing this, and how did you verify what it produced?
10. What did you decide *not* to include, and why?
11. Who, by name, stands behind the content and its facts?
12. What are the three biggest risks to this working, and your plan for each?

### And on AI:

13. Where does AI sit in your production process, and what of ours goes into it?
14. Is any of our data processed by an AI tool? Under what data processing agreement, with which sub-processors, and where is the data held?
15. Will anything we give you be used to train a model?
16. Who verifies AI-produced content, and who is accountable, by name, when it is wrong?
17. Where do you expect AI to create a material productivity benefit, and how does that assumption affect your price, timetable and quality?

**RULE OF THUMB**

A credible supplier answers proportionately, and says where an answer depends on final design, security review or contract terms. Evasive, inconsistent or unsupported answers, particularly on evaluation (7, 8 and 9) and AI (13 to 17), warrant clarification and may signal risk.



---

## Evaluation: what evidence to require

Most training is evaluated with a "happy-sheet" and nothing more. That tells you whether people enjoyed the day, which is a weak guide to whether they learned anything or changed how they work. Require more, and specify it up front.

- **Reaction** (did they rate it well?): enjoyment alone is a weak indicator; perceived usefulness tells you slightly more. Collect it, do not lean on it.
- **Learning** (did knowledge or skill change?): a pre/post check, scored.
- **Behaviour** (are they working differently?): a 30/60/90-day signal, from managers or the work itself. The level that matters most and is skipped most often.
- **Results** (did an operational metric move?): worth attempting, rarely cleanly attributable. A supplier claiming a clean financial return is overselling; an honest attempt names its limits.

Two cautions for the panel: these four are **not a chain**: a warm reaction does not predict learning, and learning does not guarantee changed behaviour. And ask for two **leading indicators** (drop-off, engagement, participation) that tell you it is working before the behaviour data lands.

---

# Accessibility, inclusion and the law you are accountable for

For a public sector body, accessibility is a legal duty, and the accountability sits with you as much as with the supplier. Specify it, and check it on delivery.

- The [Public Sector Bodies \(Websites and Mobile Applications\) Accessibility Regulations 2018](#) govern public sector websites, most intranets and extranets, and in-scope mobile apps; some content is exempt and a disproportionate-burden assessment exists, so confirm the position for your platform. Reasonable-adjustment duties arise under the [Equality Act 2010](#) in Great Britain and the Disability Discrimination Act 1995 in Northern Ireland.
- [WCAG 2.2 level AA](#) is the floor to specify: contrast of at least 4.5:1 for body text, captions on video, no meaning carried by colour alone, everything usable by keyboard.
- WCAG alone does not make learning accessible. Ask also for accessible source files, compatibility with your platforms and assistive technology (tested, not assumed), alternative formats, accessible assessment design, and remediation timescales for anything found after delivery.
- Plain English, front-loaded, acronyms defined on first use.
- Inclusive examples reflecting the real workforce, audited as a set, never tokenised.
- Impartiality: content complies with your organisation's code of conduct and presents contested issues accurately, fairly and without party-political advocacy. Use anonymised, composite or historical cases for sensitive areas.
- UK GDPR and the Data Protection Act 2018 for any learner data, with a stated retention policy.
- The commissioning body's applicable Welsh Language Standards in Wales; language provisions in Northern Ireland where relevant.

**RULE OF THUMB**

If a proposal treats accessibility as a single reassuring sentence, treat that as a red flag. Compliance is detailed or it is absent.

# Green flags and red flags at a glance

*A one-look guide for reading a proposal.*

| Look for                                        | Be wary of                                        |
|-------------------------------------------------|---------------------------------------------------|
| ✓ Objectives with verbs, conditions, standards  | ✗ Objectives full of "understand" / "be aware of" |
| ✓ Designed backwards from a named outcome       | ✗ Starts from content and topics                  |
| ✓ Fits the learner journey, with follow-through | ✗ A single event, nothing after it                |
| ✓ Live time justified minute by minute          | ✗ A day of talking at people                      |
| ✓ A lasting job aid or template                 | ✗ Slides and a certificate                        |
| ✓ Accessibility built in, in detail             | ✗ One reassuring sentence about access            |
| ✓ Behaviour-level evaluation in the bid         | ✗ Happy-sheet only, or no plan                    |
| ✓ Specific to your context                      | ✗ Generic, sector-agnostic, name-swapped          |
| ✓ Verified facts, cited sources                 | ✗ Confident, unsourced statistics                 |
| ✓ A named human accountable                     | ✗ No one's name on it                             |

---

## Sources checked

Time-sensitive claims in this guide were checked against the following sources on 5 August 2026.

- Civil service office attendance: [GOV.UK, minimum 60% expectation](#).
- Places for Growth relocation figures (reported March 2024): [GOV.UK collection](#).
- National School of Government and Public Services announcement (20 January 2026): [Government Skills](#).
- Post Office Horizon IT Inquiry, final report volume 1 (8 July 2025): [inquiry reports and statements](#).
- UK Covid-19 Inquiry module reports, including procurement (14 July 2026): [inquiry site](#).
- Ministry of Defence Afghan data breach, made public 15 July 2025: [Defence Committee inquiry](#).
- Neurodivergence estimate (around one in seven): [Local Government Association](#); an aggregated estimate, not a census figure.
- Legislation, all at [legislation.gov.uk](#): [Equality Act 2010](#); [Public Sector Bodies Accessibility Regulations 2018](#); [Data Protection Act 2018](#) and UK GDPR; [Data \(Use and Access\) Act 2025](#); [Procurement Act 2023](#).
- Standards and policy: [WCAG 2.2](#); [Government Security Classifications](#); ICO guidance on processor contracts and international transfers at [ico.org.uk](#).
- AI design-judgement research: [Dr Philippa Hardman, "AI Can Build Your Course, but Can it Design the Learning?" \(July 2026\)](#) and ["From Designing Courses to Designing Work" \(June 2026\)](#).

---

# A glossary for new commissioners

## **ADDIE**

A systematic design approach (Analyse, Design, Develop, Implement, Evaluate), often run sequentially though it can be iterative. Suits settled, high-stakes content.

## **SAM**

Successive Approximation Model. Iterative and prototype-driven; often used for fast-moving topics.

## **Backwards Design**

Designing from the outcome and its evidence first, activities last. A good sign in a proposal.

## **Bloom's taxonomy**

A ladder of cognitive verbs (remember, understand, apply, analyse, evaluate, create) used to write assessable objectives.

## **Kirkpatrick's four levels**

Reaction, learning, behaviour, results: the standard, imperfect frame for evaluation. Four questions, not a guaranteed chain.

## **Synchronous / asynchronous**

Live (a workshop, a call) versus self-paced (reading, video, e-learning).

## **Job aid**

A checklist, template or one-pager used at the point of work. Often the most valuable output.

## **WCAG 2.2 AA**

The accessibility standard your digital content must meet.

---

# A one-page commissioning checklist

## Before you go to market

- ☐ Confirmed training is the right intervention (not a process, tools or incentives problem).
- ☐ Written the outcome as a behaviour change, with a way to see it.
- ☐ Checked what already exists in your sector (Government Campus and Civil Service Curriculum, College of Policing, NHS learning platforms) so you are not duplicating.

## In the specification

- ☐ Objectives standard stated (verb, condition, standard; "understand" only with an observable demonstration).
- ☐ Design reflects the learner journey, time constraints and organisational context, with follow-through and a job aid.
- ☐ Accessibility and legal requirements (WCAG 2.2 AA, applicable accessibility regulations and equality law, UK GDPR, impartiality, applicable Welsh Language Standards).
- ☐ AI clause: transparency, verification including omission checks, data processing agreement where personal data is processed, classification handling, productivity assumptions stated.
- ☐ Evaluation plan required in the bid (behaviour level plus leading indicators).
- ☐ IP, licensing and named accountability set out.

## Evaluating bids

- ☐ Scored each proposal against the ten markers of good.
- ☐ Asked the seventeen supplier questions; noted hesitation on evaluation or vagueness on AI.
- ☐ Applied the green-flag / red-flag table, especially the AI-slop test.

## On delivery

- ☐ Accessibility checked, not assumed.

- ☐ Facts spot-checked against live sources.
- ☐ Evaluation data scheduled and owned.

**NEED A HAND?**

EnlightenWorks designs and delivers professional training. If you would like help shaping a specification, [talk to us](#).

© EnlightenWorks 2026

Share this freely within your organisation, with attribution. Written for public sector commissioners as a free reference.

Published 4 August 2026 · Last reviewed 5 August 2026 · Next review February 2027.

EnlightenWorks is independent of, and not endorsed by, any UK government or public body. Government sources referenced are publicly available on GOV.UK and were checked at the last-reviewed date. ADDIE, SAM and Kirkpatrick's Four Levels are the work of their respective originators; this document summarises them for educational purposes.