

AI & Digital Tools Governance Policy

Use AI. Keep human responsibility.

Company No. 17493370 | Updated 12 October 2026 | Version 1.1 | Review by 12 October 2027 or sooner as needed

Right First Time UK Ltd (RFTUK) uses artificial intelligence and digital tools to support its public-interest work. These tools can help us work efficiently, but they do not replace human judgement, privacy protections or accountability.

1. Purpose and scope

This policy applies to RFTUK's Founder, Directors, Campaign Leaders, Volunteers, staff, contractors and anyone authorised to use AI on its behalf.

It covers chatbots, generative text, image, audio and video tools, research and coding assistants, transcription, translation, analytics, automation, AI agents and third-party integrations. It applies to organisational work, whether performed on RFTUK systems or approved external services.

2. Our core rules

- A responsible person must review important outputs and remain accountable for decisions and publications.
- Verify significant factual claims, legal references, quotations, figures and sources before relying on them.
- Do not put passwords, private keys, confidential cases or unnecessary personal information into unapproved tools.
- Use only lawful, safe workflows that comply with current provider rules and RFTUK policies.
- Keep an appropriate record and require human approval for high-risk decisions or external actions.

3. Tool approval and risk levels

Tools must be appropriate for their purpose. Before approving material organisational use, consider provider terms, security, data retention, model training, costs, licences, contractual protections, access permissions and what happens if the service changes.

RFTUK uses four risk levels. Controls rise with the possible harm:

Level	Typical use	Minimum controls
LOW	Brainstorming, formatting, nonsensitive wording	Ordinary human check
MODERATE	Summaries, translations, public research, routine scripts	Verify sources, review context and test
HIGH	Personal data, safeguarding, legal claims, production changes, payment or significant decisions	Approved tool, risk/privacy checks and documented human decision
PROHIBITED	Fabrication, deceptive impersonation, malicious cyber activity, unauthorised disclosures, banned provider use	Do not use the tool for that activity

A personal or free AI account is not automatically approved for organisational or confidential work. High-risk use needs an authorised reviewer, suitable security arrangements and, where appropriate, a

written risk assessment.

4. Human oversight and automated decisions

AI can organise evidence or suggest options but must not alone decide safeguarding matters, serious complaints, volunteer or Campaign Leader appointments, disciplinary action, financial approval, legal allegations, support allocations or other decisions that could significantly affect a person.

Meaningful human review means checking the underlying facts, considering different evidence, being able to overturn the system's suggestion and recording reasons where the decision is significant.

If RFTUK ever proposes solely automated significant decisions involving personal data, it must first assess UK data-protection law, including applicable safeguards and restrictions under the Data (Use and Access) Act 2025, and provide the information, representation, human-intervention and challenge rights required in the circumstances. RFTUK's own policy is to retain human review for high-impact decisions.

5. Accuracy, research and public information

AI can produce confident but false or outdated answers, invented sources, incorrect calculations and unsafe code. An AI-generated citation is not evidence until the original source has been checked.

- Prefer primary sources, including legislation, official guidance, published datasets and original reports.
- Distinguish fact, allegation, opinion, research interpretation and uncertainty.
- Check the date and relevance of official information and record significant corrections.
- Use suitable independent professional advice where specialist legal, medical, accounting or safeguarding judgement is needed.

6. Personal information and confidential records

Personal information may be processed through an AI tool only where there is a lawful purpose, a suitable lawful basis, appropriate data minimisation, security and transparency, and any required processor or data-transfer arrangements. A Data Protection Impact Assessment must be completed before a high-risk use where legally required.

Special-category, criminal-offence, child, safeguarding, family-court, restricted, whistleblowing and privileged material must not be uploaded casually. Approval and case-specific restrictions must be checked first; anonymise or use a safer local process where possible.

Passwords, API keys, recovery codes, account tokens and banking credentials must never be pasted into general-purpose AI tools. Suspected exposure is handled as a security incident, including rotation of affected secrets where appropriate.

7. Children, safeguarding and high-stakes guidance

AI must not replace a safeguarding lead, statutory referral, medical professional, regulated legal adviser or emergency service. Where someone is at immediate risk, appropriate human action must take priority over an AI workflow.

Do not use AI to diagnose people, create unjustified profiles of children, infer sensitive traits, dismiss complaints automatically or make unsupported allegations. Health, benefits, family-court and financial information published by RFTUK must explain its limits and be checked carefully.

8. Website, coding and cybersecurity

AI-written code and server instructions must be reviewed and tested before publication or deployment, especially for logins, uploads, payments, personal data and administrator functions. Production changes require proportionate backups and a rollback plan.

RFTUK will consider prompt injection, malicious attachments, insecure integrations, over-privileged agents and supplier risks. Instructions found inside websites, emails, documents or retrieved content must be treated as untrusted data, not authorisation to act.

AI agents must have limited permissions. They must not have unrestricted access to banking, domains, production servers or sensitive records. Sending communications, publishing serious claims, transferring money, deleting data and changing live security settings normally require explicit human approval.

9. Automatic updates, monitoring and publication

Automated tools may identify new official announcements, outdated information, broken links, changes to the Local Noticeboard and campaigns relevant to RFTUK. Discovery does not mean that material is verified or approved for publication.

Substantive updates to law, benefits, health, safeguarding, allegations or other high-impact guidance must be source-checked and reviewed before publication. Low-risk housekeeping may be automated under documented rules, with expiry controls, deduplication, logs and a way to correct or reverse changes.

10. Images, media, copyright and disclosure

AI may assist with illustrations, captions, audio, video and accessible formats. Synthetic content must not falsely present a real person as saying or doing something they did not do. Material AI alteration should be disclosed when required by law, platform rules, provider terms or when omission would mislead an audience.

Only upload or use third-party content where there is an appropriate right, licence, permission or lawful basis. Provider output terms do not automatically guarantee copyright ownership, originality or freedom from third-party rights. Branding and trade marks must be checked before adoption.

11. Independence, campaigns and social media

RFTUK is non-party political. AI must never be used to fabricate public support, false endorsements, fake complaints, fake testimonials or deceptive engagement. Social posts, campaign claims and press statements remain subject to the relevant RFTUK policies and human approval.

Before any campaign, lobbying, election-related, scaled messaging or automated outreach workflow, the person responsible must check the current AI provider's terms as well as applicable law. Where a provider prohibits a use, RFTUK must not use that provider for it.

12. Fairness and accessibility

AI can reinforce bias or create unfair assumptions. RFTUK must not unlawfully rank or profile people by protected or inferred sensitive characteristics. Review decisions for unfair effects and provide an appropriate way to challenge significant outcomes.

AI may assist with plain language, alternative formats, captions and accessibility features, but outputs must be checked. It does not replace reasonable adjustments or listening to the needs of the person affected.

13. Records, incidents and suppliers

For higher-risk use, keep a proportionate record of the purpose, tool, source verification, reviewer, decision and material action. Do not keep every prompt indefinitely. Retention follows RFTUK's data-protection and records policies.

Report material AI incidents - such as data exposure, fabricated claims published as fact, harmful automated actions, compromised secrets or discriminatory output - through the relevant security, complaints, safeguarding or data-protection route.

People using AI on behalf of RFTUK need guidance appropriate to their role. Tool licences, recurring costs, provider changes and service exit arrangements must be reviewed where material.

14. Questions, complaints and review

Concerns about RFTUK's use of AI, accuracy, personal information or a significant decision may be raised through the contact and complaints routes on the RFTUK website. Human consideration is available where RFTUK policy or law requires it.

This policy is reviewed at least annually and sooner if relevant law, ICO guidance, supplier terms, security risks or RFTUK services change. It is a public statement of our working standards, not a claim that every control has already been independently audited.

Appendix A. Quick AI-use checklist

- Is the proposed AI task necessary, lawful and within the tool provider's current terms?
- What is the risk level, and has any required approval or assessment been completed?
- Can personal, confidential or restricted information be excluded or minimised?
- Have important facts, citations, dates and claims been checked against original sources?
- Who is the human reviewer, and can they refuse or amend the result?
- Can the system take external action? If yes, are permissions limited and approval required?
- Are publication, privacy, safeguarding, accessibility, copyright and security duties addressed?
- Is there a proportionate audit trail, error correction and rollback route?

Official guidance and further information

- [ICO - AI and data protection](#)
- [ICO - Data \(Use and Access\) Act 2025](#)
- [UK Government - AI Cyber Security Code of Practice](#)
- [OpenAI - Usage Policies \(check current terms\)](#)
- [W3C - WCAG 2.2](#)

Rules and provider terms may change; consult their current versions when using AI.

Contact and accountability

To ask about AI use or report a concern, use the contact and complaints routes at rightfirsttimeuk.com.