

Document Details: Clarification Q&A in response to the call for proposals

Challenge: **Detecting changes in authorship within online communication**

Deadline for questions: 19/09/2025

#	Question	Answer
1.	What are the organisation details that need to be provided in the phase 1 submission?	The name of the company along with a contact number, an email address, a company number, and the address of the company, would be helpful.
2.	Should phase 1 be precisely costed or is a round-number estimate sufficient?	Please confirm your costings are in line with the budget for this challenge, a breakdown of costs is not required at this stage. Please see the challenge form for details.
3.	How is benefit supposed to be elaborated - broadly with respect to possible uses of such a system, or narrowly with respect to what it should be able to do?	Without going into detail at this stage, it would be good to have both a broad understanding of uses and an idea of what it should be able to do.
4.	Is there a necessity to create new knowledge and IP or is use of existing entirely sufficient?	There is no necessity to create new IP.
5.	The licence that is granted to HMGCC - how cross government is that? Can we still sell our product to other UK gov agencies after completion?	The licence granted to the Funding Party – in either case HMGCC and DSTL, acting on behalf of the Crown – permits wider government utilisation of the Project IP solely to the extent necessary for the utilisation of the project deliverables. It does not extend to any further internal commercialisation.

		The Co-Creation terms are intended to be non-restrictive, allowing suppliers (who retain ownership of their IP) to continue exploiting and commercialising their IP after a Challenge, including with other parts of government.
6.	Will HMGCC provide any example training datasets for development and evaluation, or should applicants identify and use their own?	Applicants will have to identify and use their own training data for the purpose of this challenge.
7.	What is the upper bound on message length, and what variety of media/forms are in the input data (e.g. assessing attribution for an author across email, WhatsApp, texts)	The communication length can vary considerably, with the minimum being about two short paragraphs. Essentially an email style layout, with the possibility of cross-genre, if the solution has this capability.
8.	How will the final demonstrator be evaluated? For example, will HMGCC use a private, held-out dataset to test the solution's performance, and if so, could you share any characteristics of that dataset (e.g., languages, topics, or text lengths)?	We expect applicants to provide datasets that thoroughly test the features of their solution in line with the scope of this challenge, this will include different languages, topics and stated minimum /maximum text length.
9.	Is the explainability requirement intended to be linguistically interpretable (e.g., changes in syntax, vocabulary), algorithmically interpretable (e.g., SHAP values, feature importance), or both?	The explainability requirement may use one or multiple methods, however it needs to be understood by both specialists and non-specialists
10.	If applicants generate synthetic or scraped training data, what minimum anonymisation standard is expected to ensure compliance with UK law and GDPR?	In order to anonymise personal data so it falls outside the scope of the data protection law, the data must be effectively anonymised so that people are not (or no longer) identifiable. Anonymisation is a broad term and covers various techniques and approaches to prevent identifying people. The standard to achieve effective

		anonymisation will depend on the specific dataset. Specialist expertise and/ or legal advice may be required. For further guidance see the website of the information Commissioner's Office.
--	--	--