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

Challenge: Automated audio capture and transcription for Defence wargaming

Deadline for questions: 01/06/26

#	Question	Answer
1.	Do we specifically need to comply with any other security requirements?	There are no other security requirements other than those mentioned in the terms and conditions.
2.	The Company Number field was too small for our non-UK company number. Can we include this in our submission	We are in the process of updating the Company number field to a free text box, if your company number is too lengthy to fit, it is acceptable to include this in your submission.
3.	Are we expected to be able to evidence our ability to unilaterally develop all the challenge requirements ahead of the submission or would you be willing to accept a submission of our capability with limitations regarding infrastructure with the opportunity for you to introduce us to a preferred infrastructure supplier who we can work with during the 12-week project to enable our transcription capability to be deployed for Defence wargaming if our solution was selected?	Yes, we would be willing to accept a submission of your capability with limitations regarding infrastructure with the opportunity for us to introduce you to a preferred infrastructure supplier who you can work with during the 12-week project to enable your transcription capability to be deployed for Defence wargaming if your solution was selected. □ Full, unilateral capability is **not required at submission stage** □ Proposals may include **identified limitations**, including infrastructure gaps □ Vendors must: • Clearly define required infrastructure and dependencies • Present a **credible plan to deliver the full capability within the Proof of Concept (PoC) period** □ The Authority will **not provide pre-selected infrastructure partners**, but collaboration is encouraged.

4.	As a non UK company are we eligible to apply for this challenge.	Yes, Co-Creation welcomes applications from international organisations, provided they are not subject to *UK arms embargo restrictions*. Please note that all proposal costs must be submitted in GBP, Note:-there may be fluctuations in the exchange rate.
5.	When applying from outside the UK, are UK-based delivery partner or sub-contractor required for any element of project delivery.	There are no UK delivery partner or sub-contractor mandated for this challenge however, it would be good to indicate how you envisage delivering any physical items that you have specified in your application.
6.	What are practical considerations for a non-UK applicant regarding contract execution, payment, and on-site demonstration requirements at the TRL 6 proof-of-concept stage.	Please refer to the advertised terms and conditions for details on contract execution. For demonstrations with HMGCC Co-Creation, consider whether deliverables can be shown via video link, as face-to-face meetings are unlikely. Think about any time difference as you may need to attend meetings outside your core hours. Timings may vary depending on project needs.
7.	Are the participants in the wargames in a fixed location or do they move around the room?	It could be either. Also take into account participants will take breaks and move in and out of the secured area.
8.	Should the solution include microphones or are these already in place.	Assume not. It would be helpful if you stated any microphone requirements in your proposal.
9.	The requirement asks for speakers to be identified but also specifies that “Proposals must consider how anonymisation of Personal Identifiable Information will be managed”. What personal information are you expecting to be anonymised.	Speakers are to be identified by role, not by personal name. If needs be, an initial introduction can be made at the beginning, with personal names to be anonymised.
10.	Can we assume fixed seating/roles, or will people move between tables and rooms during the game?	It could be either. Also take into account participants will take breaks and move in and out of the secured area.
11.	Do you expect speaker identification to be by named individual, role only, or both? Can the speaker identity be manually assigned?	Speakers should be identified by Role only and a Speaker identity should be able to be manually assigned if needs be.

12.	What level of transcription accuracy is considered acceptable for the proof of concept?	"No fixed percentage is mandated at this stage; however, solutions should aim to approach near-perfect transcription accuracy in operational conditions and demonstrate clear improvement over current manual baselines." Solutions should therefore demonstrate: 1. **High-fidelity transcription of simultaneous, overlapping conversations** 2. **Accurate recognition of MOD-specific terminology and acronyms** 3. **Consistent speaker identification (diarisation) across multiple participants** 4. **Minimal reliance on post-session manual correction** 5. **Ability to review and correct transcripts where required** The emphasis is on **operational usefulness rather than raw percentage accuracy**, i.e. transcripts must be sufficiently accurate to: A. Support reliable post-exercise analysis B. Reduce or eliminate key data loss observed in manual transcription. C. Enable rapid extraction of insights and lessons. Any Solutions should also provide: **Confidence scoring or error indication mechanisms** (e.g. colour grading), and **Efficient correction workflows** to address residual inaccuracies.
13.	What is the acceptable latency between speech and live transcript availability?	"No fixed latency threshold is mandated; however, solutions should deliver transcription in near real-time with minimal perceptible delay, sufficient to support live monitoring, tagging, and correction during sessions." Q13, 27 and 36 link to latency.
14.	How many rooms or breakout areas must be captured simultaneously in the prototype?	"No specific room count is mandated; however, solutions should demonstrate the ability to capture and process multiple concurrent team discussions and show how the system would scale to full wargaming environments with multiple breakout areas," but assume a Minimum: 3
15.	Will the system be allowed to use pre-loaded glossaries, acronyms, and scenario documents?	Yes, and this will be expected to be updated per wargame.

16.	What audio infrastructure already exists in the wargaming environment, if any	Assume none and specify minimum hardware requirement
17.	Are there constraints on microphone size, visibility, cabling, wired over wireless, or placement?	"Beyond the requirement for fully wired operation with no RF-emitting components, physical microphone design (size, visibility, placement) is not constrained, provided the solution can effectively capture multi-speaker conversations in realistic wargaming environments." 1. **Effective audio capture across dynamic, multi-speaker environments**, including overlapping speech 2. **Coverage of both primary and background speakers** 3. **Scalable deployment across rooms or groups (where applicable)** 4. **Ergonomic and practical setup for wargame environments (e.g. meeting rooms, table-top exercises)** **Mandatory constraints** 1. **Wired operation only** 2. Systems must be **fully wired** (Ethernet and/or analogue/digital audio cabling are acceptable). 3. **No wireless technologies permitted during operation** **4. No RF-emitting components**, including Bluetooth or infrared. 5. **Local data handling** 6. No reliance on cloud connectivity implies **physically connected, on-premise system architectures**. These are **hard constraints driven by classification security requirements**. Vendors should therefore propose: A **wired microphone and cabling architecture** suited to multi-speaker capture A **deployment concept** (e.g. table microphones, room arrays, distributed capture points) Justification for how their approach: 1. Maintains audio quality 2. Supports diarisation and attribution 3. Remains practical within a wargaming setting.

18.	What on-premises IT environment should the prototype integrate with?	"No specific MOD IT environment is mandated for the Proof of Concept (PoC). Solutions should operate as secure, stand-alone on-premise systems, designed to integrate with classified MOD networks in future but not dependent on them for demonstration." **Fully on-premises operation** 1. All processing, storage, and model execution must occur locally (no cloud dependency). **2.Local or local-network storage only** 3. Systems must store and process data within a **self-contained or locally networked environment**. 4. **No external connectivity required** 5.Solutions should be capable of operating in **air-gapped or highly limited networks**. *(Implied by no cloud and operation constraints)* 1. **Secure-by-design architecture** 2.Must align with MOD security requirements for handling sensitive data.
19.	What are the expectations for storing, encrypting, exporting, and deleting raw audio and transcripts?	"Detailed encryption standards, retention periods, and deletion policies will be defined through MOD security accreditation processes (e.g. DPIA/RMADS). At this stage, solutions must demonstrate secure-by-design, on-premise handling of data." **Storage** **1.All data must be stored locally** (device or local network only; no cloud storage permitted). 2.Storage must support: **3.Raw audio recordings and processed transcripts** **4.Structured, searchable transcript formats** (including edited/tagged versions) 5.Systems must be capable of **retaining data for later review and analysis**. Export and Data Access. Systems must support 1. **Export of both raw and edited transcripts** (including tagged/highlighted/corrected versions) 2. **Structured and searchable outputs** suitable for downstream analysis

		3. Data export must remain within **secure, local environments** (no external/cloud transfer). Deletion and Data Lifecycle Management **1.Raw audio must be securely deleted after a defined retention period**, in accordance with: 2.A **Data Protection Impact Assessment (DPIA)** **3.** Applicable **GDPR and MOD policy** The solution should: 1. Support **controlled, auditable deletion processes** 2. Allow retention and/or continued use of processed transcript data where appropriate. **Additional Expectations** Systems should enable: 1. **Auditability of data handling actions** (storage, access, modification, deletion) **2.Controlled model improvement using approved historical data**, subject to policy and DPIA controls.
20.	Should the natural language query function work during the event, after the event, or both?	"Post-event querying is the minimum expectation; however, solutions that enable effective natural language querying during live sessions are strongly preferred due to the operational advantages they provide."
21.	What evidence will be used to judge TRL 6 success at the end of 12 weeks?	"TRL 6 is interpreted as demonstration of a prototype system in a relevant operational environment, with integrated components and evidence of performance against key requirements, rather than a fully mature or accredited capability." An evaluation rubric will be used by the evaluators across all submissions.
22.	The brief notes that HMGCC Co-Creation utilises in-house delivery managers working Agile by default and expects clear outcomes after each sprint. What is the	We do not dictate the length of sprint that the chosen solution providers run, often Co-Creation has three sprints of 4 weeks that tie up with the payment cycle (these do not all have to be equal). Each challenge is

	expected sprint duration (e.g., 2 weeks), and what is the estimated weekly or bi-weekly time commitment required from the solution provider for collaborative touchpoints, reviews, and Co-Creation workshops with MoD sponsors?	different but as a guide allow for a 2-3 hour meeting once every four week and a one hour meeting every two weeks, with as hoc involvement where necessary.
23.	An essential requirement is the ability to accurately identify domain-specific terminology and acronyms. To successfully train or build the vocabulary datasets during the 12 weeks, will the MoD provide an unclassified dataset, dictionary, or glossary (such as the Staff Officers Planning Handbook) at project kick-off, or are applicants expected to source this terminology entirely independently?	We can provide a spreadsheet with formal MOD terminology. It has approximately 28K entries. It is only for use in support of this challenge and must be destroyed on completion of the contract.
24.	The deliverable is a TRL 6 prototype demonstrated in a relevant environment. We assume the MoD will provide access to a physical wargaming facility or mock classified setup for the final TRL 6 demonstration, or must the provider simulate this relevant environment entirely within their own facilities?	MOD will provide access to a physical wargaming facility for the final TRL 6 demonstration.
25.	An essential requirement is the ability to identify speakers and attribute them to a specific role. To ensure the technical credibility of our proposal, could the Authority clarify the operational expectation for speaker attribution? Is attribution expected to be achieved via physical routing constraints (i.e., discrete, assigned hardwired microphone channels mapped to a specific role/individual) , or is the system expected to algorithmically identify speakers biometrically from voice characteristics alone within mixed audio environments?	Speaker identification is ideally by algorithm. Speakers will likely be asked to identify themselves via their role. If names are used, there will need to be a way to anonymise/remove names.
26.	Will we be required in person for the Pitch Day presentation, will this be at an MoD/HMGCC facility?	Locations vary, but to get the best out of the pitch, in person and in this case at Milton Keynes would be preferred, however you may also be able to join remotely on request. Not all applicants are always UK based.
27.	Could we clarify acceptable end-to-end latency target between speech occurring and transcription results appearing to authorised reviewers?	No fixed latency threshold is mandated; however, solutions should deliver transcription in near real-time with minimal perceptible delay, sufficient to support live monitoring, tagging, and correction during sessions.

	Given the high-concurrency requirement of supporting up to 18 simultaneous speakers across 60 total players , could the Authority define the acceptable end-to-end latency window (in seconds) between speech occurring in the room and the text rendering on an authorised reviewer's screen?	Q13, 27 and 36 link to latency.
28.	Could the sponsors provide additional guidance on the expected microphone topology within the target environment? For example: - individual wired microphones per participant, - shared table microphones, - room microphone arrays, - or a mixture of the above.	It can be a mixture of the suggestions. However, suppliers need to take into account movement within the room and between rooms, plus the need for players to take a break and move in and out of the secure areas. Refer to question 37.
29.	With reference to the requirement to deliver a full system including all hardware and software, could the Authority clarify the operational constraints and procurement mechanisms under the £60,000 threshold? To ensure our local server architecture and commercial GPU configuration align with the deployment environment , what are the maximum Space, Weight, Power, and Cooling (SWaP-C) thresholds for the on-premises hardware footprint? Additionally, should the commercial GPUs, audio interfaces, and microphone units be itemised as a direct pass-through material cost within the financial proposal, and will the physical title/ownership of this hardware transfer entirely to the MoD upon contract conclusion?	Maximum Space, Weight, Power, and Cooling (SWaP-C) thresholds would be as per a standard Network Equipment Room/Server room. Refer to Q50
30.	The essential requirements highlight both a "Natural Language Query of transcripts" and the ability for "authorised users to review and correct errors during play". Could the Authority clarify the intended intersection of these two features? Is the expectation that real-time text amendments and corrections made by reviewers must dynamically update the indexed transcript database in real time , allowing subsequent	Post-event querying is mandatory; live querying and correction workflows are strongly encouraged. Real-time synchronisation between corrections and query outputs is not mandated, but solutions demonstrating this capability will be viewed favourably.

	Natural Language Queries to immediately reflect those manual corrections?	
31.	Will the briefing call on Monday 1st June be recorded and be made available to those who are unable to attend?	No recording of the briefing call is allowed; however questions that are asked will be recorded (we advise you to use the chat function). The list will later be published in the clarifying questions document along side this challenge.
32.	During play, is the main priority accurate live capture and correction of the transcript, or do analysts also need live analysis as the exercise runs? Or is that more of an after/between-turns thing?	Main priority is accurate live capture and correction of the transcript.
33.	Would it be useful for the system to track game phases or turns against the transcript? And how are turn boundaries usually set during play - does the umpire control and announce this?	It could be helpful, but not necessary at this stage. Turns are usually detailed in the wargame events list and announced by the facilitator/umpire
34.	Could you give us a rough sense of the typical setup - number and size of rooms, breakout spaces, general acoustics - to help with microphone design?	Rooms can vary from conference rooms with approximately up to 70 pax, to meeting rooms of 20 people. Main conference room is approximately 8 x 8m in dimensions.
35.	Which on-premises systems or interfaces would the solution need to integrate with?	Standard UK power, Ethernet network, Network equipment room/server room.
36.	**Service level targets (whole system).** · Are there service level targets for the prototype - things like availability during a 12-hour session, recovery time after a fault, end-to-end latency or accuracy?	The Challenge Form does **not prescribe fixed quantitative service level targets** (e.g. specific availability percentages, recovery times, latency thresholds, or accuracy percentages) for the Proof of Concept. However, the expectation is defined in **functional and operational terms**, aligned to the Proof of Concept (PoC) nature of the activity (TRL 5–9 prototype): **1. Availability and Duration** Systems must be capable of **continuous audio capture and operation for up to 12 hours**, in line with the critical requirement for sustained wargaming sessions. Solutions should demonstrate **credible operational stability over the duration of a session**, commensurate with real-world use. **2. Fault Tolerance and Recovery** No explicit recovery-time objective is mandated.

		However, solutions should demonstrate: **Resilience to faults** where possible **Clear approaches to error handling and recovery** Ability to minimise loss of data or continuity of transcription Evidence of how faults are managed and mitigated will form part of the overall PoC assessment. **3. Latency** No fixed end-to-end latency threshold is specified. As previously stated, systems should deliver **near real-time transcription with minimal perceptible delay**, sufficient to support live monitoring, tagging, and correction. **4. Accuracy** No fixed accuracy percentage is mandated. Solutions should aim to **approach near-perfect transcription accuracy in operational conditions**, with emphasis on: Usability for analysis Handling overlapping speech Recognition of MOD terminology Reduction in manual correction burden **5. Overall Service Level Expectation** The overarching requirement is that the system is: **Operationally usable in a realistic wargaming environment**, rather than optimised to meet specific laboratory-defined service levels. Q13, 27 and 36 link to latency.
37.	**Microphone setup (E1, E2, C4).** · How many microphones do you expect, and what type (lapel, table, room array)? · Is there a preferred supplier list or a minimum quality standard?	**Number and Type of Microphones** The Challenge Form does **not prescribe a fixed number, type, or configuration of microphones**. Vendors are expected to: **Define and propose their own microphone architecture**, appropriate to meeting the operational requirements Justify how their solution enables: Capture of **multiple simultaneous speakers** (including overlapping speech) Coverage of **both primary and background contributors**

		Effective support to **speaker identification (diarisation) and attribution** Acceptable approaches may include (but are not limited to): Table microphones Room microphone arrays Distributed capture points The key requirement is that the proposed configuration is: **Technically credible**, and **Operationally suitable for realistic wargaming environments** **2. Constraints and Key Requirements** The following constraints apply: **Fully wired operation only** (Ethernet and/or analogue/digital cabling) **No RF-emitting components permitted** (including Bluetooth or infrared) Operation must be consistent with **on-premise, secure environments** Beyond these constraints, there are **no mandated requirements** regarding: Microphone size Visibility (overt vs discrete) Exact physical placement **3. Existing Infrastructure** No existing audio infrastructure should be assumed. Vendors should: Specify the **minimum viable hardware configuration** required Provide a **deployment concept** suitable for multi-speaker wargaming settings · Is there a preferred supplier list or a minimum quality standard? **4. Supplier Lists and Quality Standards** There is **no prescribed supplier list** for microphones or audio hardware No specific minimum quality standard (e.g. make/model) is mandated However, solutions must demonstrate:

		Sufficient audio quality to enable high-fidelity transcription and diarisation, and A **credible approach to achieving consistent audio capture in multi-speaker environments**
38.	**What does "as accurate as possible" mean (E3).** · Is there a target word error rate? Does the target change between single speaker and overlapping speech, or between military terms and everyday language?	The Challenge Form does **not define a fixed quantitative accuracy metric**, such as a specific Word Error Rate (WER), for the Proof of Concept. Instead, accuracy is defined in **operational and outcome-based terms**, reflecting the complexity of the wargaming environment. **1. Overall Accuracy Expectation** As previously stated, solutions should: **Aim to approach near-perfect transcription accuracy in operational conditions**, and demonstrate clear improvement over current manual baselines. The emphasis is on: **Operational usability of transcripts**, and Reduction or elimination of **data loss and misinterpretation** seen in manual capture **2. Use of formal metrics (e.g. WER)** The Authority does **not mandate a specific WER target** Vendors may choose to present performance using standard metrics (including WER), but these will be treated as **supporting evidence rather than compliance thresholds** Assessment will prioritise: **Observed performance in a representative environment**, rather than laboratory metrics alone **3. Variation by scenario (single vs overlapping speech)** The requirement explicitly recognises that transcription difficulty varies by context. As such: Solutions must demonstrate performance across: **Single-speaker conditions**, and **Multi-speaker, overlapping conversations** (which are central to the use case)

		There is **no formal requirement to meet different percentage targets by scenario**, but: Strong performance in **overlapping, multi-speaker conditions** is critical This will be weighted heavily in assessment due to operational relevance **4. Variation by language type (MOD terminology vs general speech)** Similarly: No separate numerical targets are defined for: **MOD-specific terminology**, or **Everyday conversational language** However, solutions are expected to demonstrate: **Accurate recognition of MOD-specific acronyms and vocabulary**, and Effective use of **custom dictionaries or pre-loaded glossaries** where appropriate
39.	**What does "immediately" mean for showing an utterance in the UI (E4).** · How quickly should an utterance appear on screen for the user to correct?	The Challenge Form does **not define a fixed time threshold** (e.g. in seconds) for when an utterance must appear on screen. However, "immediately" should be interpreted in **operational terms**, aligned to the requirement for near real-time usability: **1. Core expectation** Utterances should appear in the user interface with: **Minimal perceptible delay**, sufficient to allow authorised users to **review, interpret, and correct transcripts during the live event**. This aligns with the broader requirement for **near real-time transcription** supporting live monitoring, tagging, and correction workflows. **2. Practical interpretation** For the purposes of the Proof of Concept (PoC), systems should demonstrate that: Transcript text appears **quickly enough to support real-time situational awareness** Users are able to: Observe unfolding discussion

		Identify inaccuracies Apply corrections while the conversation is ongoing The expectation is therefore **functionally immediate**, rather than tied to a specific latency figure. **3. Key outcome** The determining factor will be whether: **Users can effectively interact with the transcript during live play without disruption caused by delay.** Solutions that enable smoother, lower-latency interaction will be considered more operationally effective, but no specific numerical target is mandated.
40.	**Speaker role attribution (E5).** · Should the system work out who is "Blue Team Leader" automatically (e.g. via voice), or is it fine for the user to attribute the role? · Can roles change mid-session, for example if command is handed over?	**1. Assignment of speaker roles** The expectation is that **speaker identification is conducted by role rather than named individual**. It is acceptable for **roles to be assigned manually by the user**, where required. It can be expected that speakers can announce their roles at the beginning of play to allow the system to identify them. Systems may also incorporate **automated diarisation and attribution capabilities** (e.g. via voice recognition), but this is **not mandated as a standalone requirement** The key requirement is that the system provides: **Accurate and stable attribution of contributions to roles**, and A **practical mechanism for ensuring attribution accuracy**, whether automated, manual, or a hybrid approach · Can roles change mid-session, for example if command is handed over? **2. Flexibility in role attribution** Solutions should assume that: **Roles may change during the course of a session** (e.g. command handovers, personnel changes) The system should therefore support: **Reassignment or updating of roles during live play**, where necessary

		Maintenance of **coherent attribution across the transcript**, including where changes occur **3. Operational context** Given the dynamic nature of wargaming environments: Attribution mechanisms should be **robust to changes in seating, movement, and participation** Systems should prioritise: **Usability for analysts and facilitators**, and **Transparency in how roles are applied and adjusted**
41.	**How 18 speakers and 60 players are arranged (E9).** · Are all 18 simultaneous speakers in one room, or split across breakout rooms? Please could you provide further context.	The Challenge Form defines the requirement in terms of **speaker scale and system capability**, rather than prescribing a fixed physical layout. **1. Interpretation of "18 simultaneous speakers"** The requirement to diarise **a minimum of 18 simultaneous speakers** should be understood as: The system must be capable of handling **overlapping speech from multiple active contributors at the same time**, within a single conversational context This may occur: **Within a single room** (e.g. plenary discussion or high-tempo interaction), or Across **multiple breakout groups operating concurrently** The system should not assume a single controlled speaking environment; instead, it should demonstrate robustness to **complex, overlapping dialogue patterns**. **2. Distribution across the wargame** In practice, wargaming environments typically involve: A combination of: **Breakout rooms or syndicates (teams operating in parallel)**, and **Periodic plenary sessions or cross-team discussions** Accordingly, the requirement should be interpreted as: The system must be capable of supporting **multiple concurrent conversations**, with the total number of active speakers potentially reaching or exceeding 18 at any one time across the session.

		3. Total participant context (60 players) The requirement to support **up to approximately 60 enrolled speakers** relates to: The total number of identifiable participants across the session Not all participants will be speaking simultaneously The system should therefore demonstrate: **Stable diarisation across a large participant set**, and Ability to maintain **identity (role-based attribution) over time**, even as speakers join, leave, or change roles **4. Key design implication** Solutions should be capable of: Handling: **Dense conversational overlap within groups**, and **Parallel conversations across groups** Providing: Clear separation of discussions Reliable attribution at scale The emphasis remains on **functional capability and scalability**, rather than a prescribed physical configuration.
42.	**Tagging granularity and schema (E11, D4).** · At what level should tagging work - per utterance, per turn, per topic or per session phase? · What categories matter (team, phase, intent, classification, free text)? · Are tags chosen from a fixed list defined at session setup, or added freely by the operator during the session?	**1. Tagging granularity** The Challenge Form does **not mandate a single fixed level of tagging granularity** (e.g. per utterance, per turn, per topic, or per phase). However, the system is expected to support: **Time-stamped, segment-level tagging** of transcripts The ability to apply tags **both during live transcription and post-event review** Sufficient granularity to enable **meaningful analysis of discussions, decisions, and transitions** In practice, this implies that tagging should be flexible enough to support: Fine-grained tagging (e.g. at utterance or short segment level), and Broader structuring (e.g. topic, phase, or discussion segment), where appropriate

		The emphasis is on **analytical usefulness rather than a prescribed tagging level**. · What categories matter (team, phase, intent, classification, free text)? **2. Tag categories** The Challenge Form does not define a fixed taxonomy of tag categories. However, it highlights use cases where tagging should support: **Game turns** **Injects** **Adjudication decisions** **Phase transitions** In addition, solutions should be capable of supporting categories such as: Role or team attribution Temporal segmentation (e.g. phases or stages of play) Analytical markers (e.g. decision points, key discussions) Vendors may propose additional categories where they improve analytical utility. · Are tags chosen from a fixed list defined at session setup, or added freely by the operator during the session? **3. Tag schema flexibility** The expectation is that tagging should be **customisable and user-driven**, rather than fixed. Specifically: Systems should allow: **Pre-defined tags** established at session setup, and **Dynamic tagging during the session**, added by authorised users as required Tagging should support: Both **live application** (during play), and **Post-event refinement and editing** This aligns with the broader requirement for: **Customisable labelling**, and The ability to **review and correct transcript data**

		4. Key expectation The overarching requirement is that tagging: **Enables structured, searchable, and semantically meaningful analysis of wargame data**, rather than adhering to a rigid predefined schema. Solutions should therefore demonstrate: Flexibility Usability for analysts Alignment with wargaming processes
43.	**One admin role or multiple roles (D3).** · For the prototype, is a single administrator role enough, or do you need separate roles for example, analyst (read only) and admin (full access) roles? · Do we need to record who corrected which transcript segment and when?	**1. User roles within the prototype** The Challenge Form does **not mandate a fixed user role model** (e.g. defined separation between administrator, analyst, read-only users, etc.) for the Proof of Concept (PoC). For the purposes of the PoC: A **single administrator or operator role is sufficient**, provided it enables: Review and correction of transcripts Tagging and annotation Management of system outputs However, solutions that support **multiple user roles** (e.g. analyst vs administrator) may be beneficial, particularly where they: Improve usability in a collaborative environment Reflect realistic operational workflows (e.g. facilitators, analysts, observers) This functionality is considered **desirable rather than essential** at this stage. · Do we need to record who corrected which transcript segment and when? **2. Auditability of corrections** The Challenge Form does not explicitly mandate a requirement to record: Who made a specific correction, or When that correction occurred

		However, the system is expected to support **secure and auditable data handling**, including the ability to: Review and amend transcripts Maintain confidence in the integrity of the record Accordingly: Solutions that provide **audit trails for edits (e.g. user, timestamp, change history)** would be considered **strongly desirable**, as they support: Transparency of analysis Traceability of changes Alignment with future security and accreditation requirements
44.	**What does "natural language query" mean (E10).** · How advance does this system need to be? i.e. are you envisioning using an LLM (artificial intelligence) or is this pure keyword search? · If we use AI models, are there any restrictions beyond the existing no-internet and no-cloud constraints?	**1. Definition of "Natural Language Query"** The requirement for a natural language query capability should be understood as: The ability for users to interrogate transcripts using **plain language inputs**, rather than requiring structured query syntax or predefined filters. This is intended to support: Efficient interrogation of large, multi-speaker transcripts Extraction of decisions, themes, and key discussions Reduction in analyst burden during post-exercise analysis **2. Required level of sophistication** The Challenge Form does **not prescribe a specific technical implementation** (e.g. LLM-based vs keyword-based systems). Both approaches may be acceptable, provided they: Enable **intuitive querying by non-technical users**, and Deliver **meaningful, context-relevant results from transcripts** However: A **simple keyword search capability alone may be insufficient** if it does not enable effective interrogation of complex discussions Solutions that provide more **context-aware querying capability** (e.g. semantic understanding of queries) are likely to offer greater operational value The emphasis is therefore on:

		Analytical effectiveness, rather than the specific technology used to achieve it **3. Use of AI / LLM-based approaches** The Challenge Form explicitly allows for AI and LLM-enabled solutions, subject to the overarching system constraints. If AI models are used, the following constraints apply: **Fully on-premises operation** All models must run locally (no cloud-based inference or processing) **No external connectivity** Systems must be capable of operating in **air-gapped or highly limited environments** **No cloud storage or processing** All data handling must remain within the local system or local network **Secure-by-design operation at higher classification levels** Models must align with MOD security requirements and data handling policies **Controlled model improvement** Any learning or fine-tuning must use **MOD-provided or MOD-approved datasets**, subject to governance (e.g. DPIA controls) **4. Live vs post-event use** As previously stated: **Post-event querying is the minimum requirement** Solutions that support **querying during live sessions** are **strongly preferred**, but not mandatory . If we use AI models, are there any restrictions beyond the existing no-internet and no-cloud constraints? LLMs sourced from countries on the *UK arms embargo restrictions* will not be suitable.

45.	**Wargame data for training and testing (E6, D2).** · Can you share representative transcripts, glossaries or acronym lists for use during the project?	**Potential provision during the project** During the Proof of Concept (PoC) phase, there may be **opportunities to use MOD-provided or MOD-approved datasets** for: Testing Validation Controlled model improvement Any such provision would be: **Limited in scope**, and Managed in accordance with **security, GDPR, and DPIA considerations** **Key expectation** The expectation is that vendors: Demonstrate a credible approach to handling domain-specific language, including how their solution can adapt to MOD terminology with or without pre-supplied datasets.
46.	How many rooms tend to operate at the same time during war games?	1 x main 'conference room' of 60 pax, up to 10 breakout rooms between 4-20 pax
47.	How many orgs can be awarded for the proposal?	Typically 5 are invited to the Pitch Day and 1-2 awarded a contract.
48.	Does all hardware need to be COTS	The Challenge Form does **not mandate that all hardware must be Commercial Off-The-Shelf (COTS)**. However, the following considerations apply: **1. Acceptability of hardware types** • Solutions may utilise: ○ **COTS components**, ○ **Modified COTS**, or ○ **Custom-developed hardware**, provided they meet the overall system requirements. However, if hardware is bespoke/custom then the Total Cost of Ownership of provision, maintenance and upgradability will be taken into account as there is an assumption that bespoke solutions will have higher TCO.
49.	Are there any historical audio datasets with transcriptions available for testing and training a solution?	No. DEWH intends to record several wargame demonstrations from its Wargame Design course which will provide several baseline recordings. They will not be available prior to the pitch day.

50.	Should the commercial GPUs, audio interfaces, and microphone units be itemised as a direct pass-through material cost within the proposal? And will the physical title/ownership of this hardware transfer entirely to the MoD upon contract conclusion?	It is intended that the final prototype (MVP) will be handed over to the sponsor on completion for further testing. Therefore the hardware would transfer to the sponsor on completion. As mentioned in Q55, proposals can be based on a simpler demonstration (e.g. single room demonstration) provided that a clear and credible explanation of how the solution would scale to a multi-room environment. This could potentially limit the hardware procurement requirement.
51.	Can you clarify the point on Cyber Essentials Plus - is that a pre-requisite for project kick off?	Yes as stated in the Challenge Form.
52.	Moving forwards, are the reviewers part of the MOD or will they be part of the provider's team?	The reviewers of the proposals will be staff from HMGCC Co-Creation and also the sponsor (MOD - jHUB).
53.	Can you outline the anticipated position in relation to IP issues for any software provided	As per the HMGCC Co-Creation terms and conditions, project IP shall belong exclusively to the solution provider, granting the Authority a non-exclusive, royalty free licence.
54.	Is the Audio, Microphone AV system to be included in this proposal or is it separate (i.e. there will be an AV system we can plug into	Audio and Microphone AV system is to be included in the proposal. Suppliers are not to assume any AV/audio to plug into. There is a secure ethernet network and Network Equipment Room (NER) / Server room.
55.	As the multi-room use case will push up the hardware (microphone) cost and therefore limit the budget for innovation - can the submitted proposal be based on a single room with a clear explanation of how the multi-room scenario would be implemented.	Yes, it is acceptable for proposals to be based on a **single-room demonstration**, provided that this is supported by a **clear and credible explanation of how the solution would scale to a multi-room environment**. **1. Expectation at Proof of Concept (PoC) stage** The Challenge Form does not mandate that vendors must physically demonstrate a full multi-room deployment within the Proof of Concept. Instead, the expectation is that: • Vendors demonstrate **core system capability in a representative environment** (e.g. a single room), and

			• Provide a **credible and technically coherent approach to scaling** that capability to:
			  ○ Multiple concurrent teams, and
			  ○ A wider wargaming environment

			2. Multi-room requirement interpretation
			
			As previously stated:
			
			• The requirement for multiple rooms/breakout areas is **capability- and scalability-driven**, rather than a fixed demonstration constraint
			• Solutions must ultimately be capable of supporting:
			  ○ **Parallel conversations across multiple groups**, and
			  ○ Aggregate participant and speaker scales

			3. Proposal guidance
			
			Where proposing a single-room Proof of Concept (PoC), vendors should:
			
			• Clearly define:
			  ○ The **baseline configuration** demonstrated
			  ○ The **assumptions and limitations** of the single-room setup
			• Provide a **scaling concept**, including:
			  ○ How additional rooms or groups would be supported

		○ Required changes to: ▪ Hardware (e.g. microphones, capture nodes) ▪ Processing architecture ▪ Data handling and integration • Demonstrate that the solution architecture is: ○ **Modular**, and ○ **Capable of extension without fundamental redesign**
		4. Assessment approach Solutions will be assessed based on: • Demonstrated performance in the Proof of Concept (PoC) environment, and • The **credibility of the proposed scaling approach**, rather than requiring full physical replication of all target conditions
56.	Question on speaker identification: are we looking for single player per mic or many players per mic? Are players identifying themselves by callsign or is it intended that they are identified by voice signature?	**1. One speaker per microphone vs multiple speakers per microphone** The Challenge Form does **not mandate a specific microphone-to-speaker ratio** (i.e. one speaker per microphone). Solutions may adopt either approach: • **Single speaker per microphone**, or

		• **Multiple speakers per microphone** (e.g. table or room-based capture) However, vendors are expected to propose a configuration that can: • Support **multi-speaker environments with overlapping speech** • Enable **accurate diarisation and attribution at scale** (up to 18 simultaneous speakers, ~60 participants per session) In practice, this means that: The microphone configuration must be justified as capable of delivering the required transcription and speaker attribution performance, regardless of the number of speakers per device.
		2. Method of speaker identification (role attribution) The expectation is that: • **Speaker identification is by role rather than named individual** • The system should support **stable attribution of contributions to roles across the session** With regard to how this is achieved: • **Manual or assisted role assignment is acceptable and expected where required**

		- For example, initial identification at the start of a session or operator-led assignment - Fully automated identification based solely on **voice signature is not mandated**, although it may be used as part of a solution
		3. Practical operational approach A typical and acceptable approach may include: - An **initial introduction or calibration step**, where participants identify themselves - Mapping of speakers to **roles (not personal identity)** - Use of: **Diarisation (voice separation)**, and - Optional **voice-based recognition to support attribution**, supplemented by manual correction where required Solutions should assume: - Participants may **move between locations** - Roles may **change during the session** - Attribution must therefore be **maintainable and adjustable in real time**
57.	In terms of Ethernet Cabling being available, is this just structured basic cabling, or does this cabling terminate at an Ethernet switch that will be available for use ?	Both.

58.	What mics do you already have in place? Can these be utilised?	Assume no microphones are in place. If necessary, specify a standard to be achieved.
59.	How frequent are players moving between rooms and around a room?	Players may be required to move to breakout rooms and back into the main conference room. Players will also need to move out of the secure area for breaks.
60.	Further to a previous question - can we then assume that participants will be stationary (and can have wired microphones) or do people have to be able to move around.	Players may be required to move to breakout rooms and back into the main conference room. Players will also need to move out of the secure area for breaks.
61.	Will the SME get to keep the IP?	As per the HMGCC Co-Creation terms and conditions, project IP shall belong exclusively to the solution provider, granting the Authority a non-exclusive, royalty free licence.
62.	Can the SME export to the US post project?	As per the HMGCC Co-Creation terms and conditions, project IP shall belong exclusively to the solution provider, granting the Authority a non-exclusive, royalty free licence. Therefore the Solution Provider can export if they wish.
63.	Are discussions typically one person at a time in the conversation, or will groups often have multiple overlapping speakers at once?	Refer to Q41
64.	How many *spaces/rooms* in the example use case Are players moving around or static? Size of the conference room? Max total occupants of any room at any one time	Partially answered in Q41. 1 x main 'conference room' of 60 pax, up to 10 breakout rooms between 4-20 pax. Players may be moving from conference room to breakout rooms and back again. Players will also be moving in and out of secure areas for breaks. Conference room size: 8 x 8 M Max total in conference room: 70
65.	How many tables (with how many people) in the conference room?	~10 tables, approximately 60 people.
66.	If after the 12 week prototype development the capability needs to be capable of delivering a higher classification level example of the solution that includes microphones.	Assurance for higher classification levels will begin at the end of the prototype phase should the prototype prove to be a viable solution. The

	Is higher classification security assurance ongoing or at the end or our responsibility	assurance will be conducted by the MOD and will involve the supplier's co-operation in regard to testing and verification.

67.	You mentioned wired only solutions. Could you confirm does this mean lapel mics are excluded by definition. Or is there a work around for lapel mics?	**1. Interpretation of "wired-only" constraint**
		
		The requirement for **fully wired operation with no RF-emitting components** is a **hard constraint** driven by security requirements (in higher level classification environments).
		
		This means:
		
		• **Wireless transmission is not permitted under any circumstances**, including:
		  ○ Bluetooth
		  ○ Wi-Fi
		  ○ Standard wireless lapel microphone systems
		  ○ Any RF-based bodypack transmitters
		
		Accordingly:
		
		Conventional wireless lapel microphones are not permitted.
		
		2. Use of lapel microphones (wired options)
		
		Lapel microphones are **not excluded by definition**, provided they are implemented in a **fully wired configuration**.

			Acceptable approaches may include:
			
			• **Wired lavalier (lapel) microphones**, physically connected via cable to:
			  ○ Recording/processing devices, or
			  ○ Local audio interfaces
			
			However, vendors should consider:
			
			• Practical constraints associated with:
			  ○ Participant movement
			  ○ Cable management in dynamic environments
			  ○ Scalability across multiple participants
			  ○ On body secure storage and management of secure storage devices.

			3. Practical expectation
			
			The Authority does not mandate a specific microphone type; however, any proposed solution must:
			
			• Comply fully with the **no RF / fully wired constraint**
			• Be **operationally practical in a wargaming environment**
			• Support:
			  ○ Multi-speaker capture
			  ○ Overlapping conversations
			  ○ Reliable diarisation and attribution

		Vendors are therefore expected to:
		
		• Propose a **coherent and practical wired audio architecture**
		• Justify how their approach balances:
		  ○ Audio quality
		  ○ Usability
		  ○ Scalability

68.	Is the portability for use as a deployable system or use on fixed sites	**1. Intended interpretation** Portability should be understood as: The ability to deploy the system **across different locations and environments**, rather than being permanently installed in a single fixed site. This may include both: • **Relocation between fixed facilities** (e.g. different rooms, buildings, or sites), and • Potential use in more **deployable or temporary setups**, where appropriate **2. Proof of Concept (PoC) expectation** For the Proof of Concept (PoC):

			• There is **no requirement to demonstrate a fully deployable (field-based) system**
			• Solutions may be:
			  ○ **Fixed for demonstration purposes**, or
			  ○ Designed with portability in mind but not fully exercised during the PoC

			3. Practical considerations Where portability is addressed, vendors should consider: • Ease of **setup, teardown, and reconfiguration** • Transportability of hardware components • Ability to operate in **varying room layouts and environments** However, these considerations should be balanced against: • Core functional performance (e.g. transcription, diarisation) • Security constraints (wired, no RF, on-premise operation)
			Summary Position • Portability is **desirable but not mandatory** • It is intended to support **use across different locations**, not solely fixed-site deployment

		• A fully deployable system is **not required at Proof of Concept (PoC) stage**

69.	Roughly how many reviewers do you envisage interacting with the system to ensure transcription accuracy during play	**1. Expected approach at Proof of Concept (PoC) stage**
		
		For the purposes of the prototype:
		
		• A **small number of authorised users (e.g. one or more analysts/facilitators)** is sufficient to:
		  ○ Review transcripts
		  ○ Apply corrections
		  ○ Conduct tagging and annotation
		
		There is **no requirement to demonstrate large-scale concurrent reviewer activity** during the PoC.

		2. Operational context
		
		In typical wargaming settings:
		
		• Transcript review and correction is likely to be conducted by:
		  ○ **A limited number of analysts or data capture personnel**, rather than all participants
		• These individuals would:
		  ○ Monitor live transcripts
		  ○ Intervene selectively to correct errors or apply tags

			The system should therefore support: • **Low user concurrency**, with the ability to scale if required • Efficient workflows that minimise the need for constant manual intervention
			3. Key expectation The emphasis is not on the number of reviewers, but on whether the system: **Reduces the burden on human reviewers while still allowing effective oversight and correction where required.**

			Summary Position
			
			• No fixed number of reviewers is mandated
			• A **small number of authorised users is sufficient for the Proof of Concept (PoC)**
			• Systems should support:
			  ○ Live review and correction
			  ○ Efficient workflows for a limited analyst cohort
			• Scalability to additional users may be beneficial, but is not required at this stage