

Document Details: Clarification Questions – Mobile Mimicry

Deadline for questions: Friday 28 September 2026

Publish date: Tuesday 6 October 2026

1.0 Project/Clerical

	Question	Answer
1.1	Registration: The challenge guidance indicates that sole innovators are eligible to apply. However, the online application form currently requires both “Organisation Name” and “Registration Number”, and the Registration Number field requires 6–10 alphanumeric characters. I do not have a registered company or organisation. Could you please confirm what a sole innovator should enter in these two required fields?	For a sole trader, you can enter your Full Name as the Organisational Name and zero’s in the Registration Number field.
1.2	Payment profile: For an SME delivering the 12-week project, can the work be structured around regular delivery milestones with corresponding invoices, for example monthly milestones, or is payment normally concentrated towards final delivery?	The project is structured around incremental payment milestones, general working towards 3 to 4 week sprints with a deliverable at the end of each Sprint.
1.3	Contracting: Can you please confirm who we will be contracting with and who we will raise the invoice to. Will	The contract will be with Cranfield University on behalf of HMGCC. See T&C’s https://hmgcc.gov.uk/terms-and-conditions

Official

	it be HMGCC or Cranfield University. We just need to clarify this for our costing system set up.	
1.4	Eligibility and contracting. Our contracting entity is registered in Estonia, with engineering teams in Estonia and Ukraine. Can you confirm that such an organisation is eligible, and is there any requirement for a UK-registered entity or UK-based personnel for contracting via Cranfield University? For non-UK staff, what would be considered equivalent to BPSS checks?	We welcome applications from any non-embargoed country. BPSS checks may be performed on non-UK staff.

2.0 Technical

	Question	Answer
2.1	Control boundary: For the minimum TRL 5 demonstrator, is it acceptable for the AI reasoning/orchestration layer to run on an on-premises workstation controlling Android emulators/devices, provided the phones execute the behaviours autonomously, or is on-device inference/orchestration itself an essential requirement?	An on-premises workstation controlling multiple devices would be preferable.

Official

2.2	Behaviour realism: Is the primary evaluation goal broad functional coverage (SMS, apps, calls, browsing, maps, etc.) or statistical realism of longitudinal human behaviour (session timing, pauses, app-switching, routines and persona consistency)? This materially changes the architecture and validation plan.	'Functional coverage' is the requirement here now. (Behavioural imitation would be a stretch, or future ambition).
2.3	Explainability: would a machine-readable action/evidence ledger — intended to replay why the agent chose each action, what it observed, and whether the action succeeded — be considered useful evidence for the guardrails/explainability requirement, or are you expecting a different form of explainability?	Explainability is for audit / compliance purposes – this should suffice.
2.4	On-device vs on-premises: Must AI inference/planning execute on the mobile device itself, or would an on-premises orchestration service controlling agents running on individual devices/emulators satisfy the requirement?	As 2.1
2.5	Operating System: Would an Android-only TRL 5 demonstrator be acceptable for this challenge, provided all essential functional requirements are met?	Yes
2.6	Device automation interfaces: Are Android Accessibility APIs, UIAutomator/ADB, Appium, or comparable automation interfaces acceptable mechanisms for interacting with applications and the operating system, or is the expectation that the agent uses only user-visible interaction methods?	Android Accessibility APIs, UIAutomator/ADB, Appium, or comparable automation interfaces are acceptable.

2.7	SMS, messaging and voice calls: For the TRL 5 demonstrator, must SMS, third-party messaging applications and voice calls all operate across live telecom networks, or can some elements be demonstrated using controlled/emulated endpoints where appropriate?	Some elements can be demonstrated using controlled/emulated endpoints where appropriate.
2.8	Messaging applications: Are there specific third-party messaging applications that the Authority expects the demonstrator to support?	The aim is to achieve a 'realistic' device configuration – we invite you to propose your own interpretation of this.
2.9	Unsupervised in training: Could you clarify what is intended by the requirement that the agent runs “unsupervised in training”? Does this refer to model training, behavioural/persona learning, autonomous execution during testing, or another process?	Autonomous execution during test, the device displaying human like activity. The goal is to have the system that requires minimal manual intervention.
2.10	External/cloud AI services: Given the on-premises requirement, are externally hosted AI inference APIs prohibited for the final demonstrator, or would a hybrid architecture be acceptable provided sensitive test data and device control remain on-premises?	Device control should remain on-premises.
2.11	Human-like behaviour: Are there defined metrics or acceptance criteria for assessing whether timing and behaviour are sufficiently human-like, or should solution providers propose their own validation methodology?	Solution providers should propose their own validation methodology.
2.12	Physical handset: The challenge describes operation on a physical phone as desirable. Would demonstrating the same agent architecture on both an Android emulator and at least one physical Android handset materially strengthen compliance with this desirable requirement?	Ideally this would use multiple physical Android handsets.

2.13	Agent location: Does the requirement for an AI agent operating on a mobile device or virtual surrogate require the model or agent runtime itself to execute within the mobile operating system, or is an on-premises agent permitted to observe and control an Android emulator or device through an appropriate control interface?	An on-premises agent permitted to observe and control an Android device.
2.14	Operational autonomy: The challenge states that “the agent runs unsupervised in training; a human can intervene during operational use.” Could you clarify whether “training” here refers to autonomous test execution, or whether applicants are expected to perform unsupervised model training as part of the demonstrator?	The reference here is autonomous execution.
2.15	Training and personas: For acceptance purposes, what does “runs unsupervised in training” mean? Is model-weight training or fine-tuning required, and how will the pre-trained personas be assessed?	The goal here is to be able to run a mobile device, unsupervised, over a period of time, that mimics the behaviour of a human user.
2.16	Deployment architecture: May model inference and orchestration run on a shared on-premises host controlling the mobile emulators, or must these components execute within each mobile instance?	Ideally the model would run on multiple mobile devices.
2.17	Platform coverage: Which mobile operating systems must the demonstrator support? Is demonstrating the requirements on a single operating system sufficient?	Yes, although a stretch, or future goal would be to run over multiple operating systems.

2.18	Messaging and telephony: For the emulated demonstration, are SMS and voice calls between emulators acceptable, or must these connect to live services? What constitutes successful voice-call placement, and are there particular messaging apps that must be supported?	Ideally, SMS and voice calls would be via live service, however, emulators would be acceptable.
2.19	Scale and validation: Is there a minimum number of concurrently active mobile instances or selectable personas expected? What criteria will be used to assess realistic human behaviour and successful TRL 5 demonstration?	At least two devices.
2.20	Technical approach: For the essential requirements, would applicants be expected to already have these capabilities at the point of application, or is it acceptable to demonstrate a credible technical approach for developing and integrating them during the 12-week project?	It would be acceptable to demonstrate a credible technical approach for developing and integrating them during the 12-week project.
2.21	Mandatory requirements: How are the desirable requirements considered relative to the essential requirements, particularly running the solution on a physical mobile device?	Running the solution on multiple physical mobile devices is the ultimate aim.
2.22	Previous tech: Do applicants need to have previously developed an AI agent specifically for mobile devices, or would you consider organisations with relevant and transferable AI and software development experience?	Relevant and transferable AI and software development experience will be considered.

Official

2.23	If an applicant has strong capability in some areas but would need additional expertise for certain elements, such as mobile implementation, would bringing in a technical partner or collaborator be considered appropriate?	Yes.
2.24	Execution architecture: For the essential requirement to demonstrate an AI agent operating in an emulated environment, is it acceptable for the behavioural and/or AI inference service to execute on an on-premises host which controls one or more emulated mobile-device instances? Or is the expectation that the complete agent/model itself executes within each emulated or physical mobile-device instance?	An AI inference service that executes on an on-premises host which controls more than one mobile-device instance is preferable.
2.25	Evaluation of human-like behaviour: Does the challenge owner have an existing evaluation methodology, reference dataset, benchmark or acceptance criteria for determining whether generated persona behaviour is sufficiently representative of natural human mobile-device usage? If not, are applicants expected to propose their own quantitative and/or qualitative behavioural-validation methodology?	Applicants can propose their own quantitative and/or qualitative behavioural-validation methodology.

2.26	SMS, messaging and voice-call demonstration: For the TRL 5 demonstrator, may the essential SMS, messaging-app and voice-call capabilities be demonstrated between controlled/emulated devices and controlled services within the on-premises test environment? Or is integration with external telecommunications infrastructure expected as part of the demonstrator?	You can demonstrate in an on-premises test environment. Integration with external telecommunications is preferred.
2.27	Architecture: Would an architecture be considered within scope where the selectable behavioural personas, longitudinal state, decision logic, constraints and explainability are managed by a persistent workload/persona engine, while mobile GUI interaction is performed by a replaceable AI interaction executor? The intention would be for the combined system to autonomously reproduce persona-dependent mobile usage rather than merely execute predetermined test scripts.	Yes, your described intent, and its underlying architecture is within scope here.
2.28	Operator technical level: How non-technical is the target user? For example, can the operator be expected to enable Developer Options on the device and follow a few guided setup steps during installation?	The user will be 'non-technical', but they will be able to follow guided setup steps during installation.
2.29	Access rights: How broad can the application's access rights on the test devices be? For example, may the agent run with root access or as a privileged system	Assume standard user-level installation.



Official

	application, or should it assume a standard user-level installation?	
--	--	--