

Response to Ofgem AI Technical Sandbox Consultation

Respondent: Sagittal AI (Sagittal Limited, Companies House #15168318)

Contact: Michael Smith, CEO, michael@sagittal.ai

Date: 20 March 2026

Submitted to: AlPolicy@ofgem.gov.uk

About Sagittal AI

[Sagittal AI](#) builds [Neo](#), a Process-Controlled AI platform for enterprise software development lifecycle (SDLC) automation. Neo has been deployed in production across regulated industries, most significantly at Telefónica, where it was used by a division of approximately 300 engineers. No security breach or material security incident has occurred in any production deployment or trial. Neo has passed independent security and legal reviews at multiple organisations in security-sensitive sectors, including banking and utilities - sectors directly analogous to those within Ofgem's regulatory remit.

We are a participant in the NIST NCCoE DevSecOps consortium under a formal Cooperative Research and Development Agreement (CRADA), alongside Google, Microsoft, IBM, GitLab, and others. In March 2026 we submitted a [formal response](#) to [NIST RFI 2026-00206](#) on the security considerations for AI agents.

Michael Smith, CEO & Co-founder brings over 25 years of engineering and product leadership across Google, Amazon, Yahoo, and SwiftKey. He has presented Neo's architecture at NCCoE events and leads Sagittal's engagement with US federal standards development, and has experience in secure software development in telecoms.

Jose Palazon, CTO & Co-founder brings over 25 years of technology leadership, including more than a decade in CTO roles. He served as CTO of Telefónica's engineering organisation - the same division in which Neo was deployed in production. He is a published author on secure programming and has presented at DEF CON, Black Hat, and ShmooCon. His security experience is directly relevant to the subject matter of this consultation.

Commercial interest disclosure: Sagittal has a direct commercial interest in the outcomes of this consultation. We would benefit from being admitted as a Named Technology Partner in the sandbox and from Neo being evaluated as part of any sandbox work on secure AI-assisted

software development. We have set out our policy arguments as candidly as we can, and we have tried to identify where they coincide with our commercial interests rather than obscure this. We ask that Ofgem assess them on their merits.

Opening Position: The Case for Expanding the Sandbox Scope

The Threat Is Real, Present, and Growing

Critical national infrastructure, and energy infrastructure in particular, is a primary target of actors seeking to disrupt the UK economy. This is not a theoretical risk. In December 2025, the Russia-aligned Sandworm group deployed purpose-built data-wiping malware against a Polish energy company, targeting operational technology during peak winter demand with the explicit goal of physical service disruption. That same year, the Russia-linked ransomware group Qilin claimed breaches of multiple US electric cooperatives, exposing sensitive operational data and demonstrating that the threat extends well beyond elite state actors to well-resourced criminal groups operating at scale. The pattern is consistent and escalating: energy infrastructure is a primary target, software is the vector, and the actors pursuing it range from nation-state operations to criminal enterprises with comparable capabilities.

What is new - and makes the present moment different - is the role of AI on both sides of this equation. CNI operators now routinely build and maintain software using AI coding tools. These tools offer genuine productivity benefits, and their adoption is accelerating. But not all AI coding tools carry the same risk profile. The first generation - autocomplete and suggestion tools such as GitHub Copilot - generates code for a human to review and execute. The emerging generation is qualitatively different: agentic AI tools reason, plan, and act autonomously, directly invoking shell commands, file systems, version control, external APIs, and infrastructure - without a human in each loop. A developer using an agentic coding tool may instruct it to "implement this feature and run the tests"; the tool may write the code, execute it, modify configuration, call external services, and commit changes, all under the developer's identity and credentials. This is not a faster autocomplete. It is a different mode of software production entirely, with a correspondingly different threat surface.

At the same time, adversaries are using AI to craft more sophisticated, faster, and more targeted attacks. The result is a structural asymmetry: AI lowers the cost and raises the sophistication of offensive operations faster than it raises defensive capability - that is, unless defenders adopt structured, auditable approaches to AI use. Without deliberate intervention, Ofgem's licensees will face a greater attack surface with fewer effective defensive measures than at any point in the sector's history.

Energy Must Own This - Not Outsource It

Ofgem is right to situate the AI Technical Sandbox within the energy sector specifically. The temptation to defer CNI software security to economy-wide bodies such as DSIT, NCSC, or others with broader remits, would be understandable but wrong.

Because energy infrastructure is a target, it is also subject to a regulatory framework - including the NIS Regulations 2018 - that already requires operators of essential services to manage network and information systems security to an outcome-based standard. What that framework does not yet address is how AI-assisted software development interacts with those obligations. That gap is Ofgem's to fill. No economy-wide body has both the sector-specific depth and the regulatory authority to do so.

Cross-sector expertise - from cybersecurity, AI architecture, and software engineering - is essential and must be drawn in. But energy has a responsibility and ability to lead CNI, and to own the outcomes.

The Proposed Use Case

We submit this response with a specific use case in mind, and invite Ofgem to consider it for a dedicated sandbox thread:

How should CNI operators use AI to develop more secure software, and adopt a defensive AI posture against AI-enabled attackers?

This use case has four concrete objectives:

1. **Demonstrate additional security** through practical, deployable approaches, not theoretical frameworks
2. **Create energy-specific standards and compliance recommendations** as far left in the development process as possible, catching vulnerabilities at the point of code generation rather than post-deployment
3. **Help companies take specific, actionable steps** against those recommendations, with measurable outcomes
4. **Show demonstrable improvement** to the defensive posture of CNI operators and UK national security as a whole

Q1: Eligibility and Participation

Do you agree with the proposed eligibility criteria for lead Participants (licensees, market participants, and operators of essential services) and the encouragement of partnerships with

technology providers, academia, and other innovators? Please explain your reasoning.

We Support the Lead Applicant Framework

The restriction of lead applicants to licensees, market participants, and operators of essential services is pragmatically sound, and we support it. Licensees and OES bring what no technology partner can fully replicate: direct access to operational data, regulatory accountability, and the sector context that makes sandbox findings meaningful rather than hypothetical. This is not primarily a technology evaluation exercise - it is an exercise in understanding how new technologies interact with the real operational, compliance, and risk environment that energy operators inhabit, and that environment is best understood by those who work within it.

We also recognise the rationale for ensuring that regulatory accountability attaches to the lead participant. Where sandbox findings have implications for live deployment, there must be a clear line of accountability. The lead applicant framework provides that, and we do not propose to change it.

The Current Framing of Technology Partners Is Too Restrictive

We have a significant concern, however, about how technology partners are positioned within the proposed structure.

As currently framed in Section 3.2 of the consultation, technology partners are delivery resources. They provide expertise, technology, or resources to support the lead applicant, but they have no governance rights, no formal input to the Steering Group, and no direct relationship with Ofgem. The lead applicant mediates all of this.

For many use cases, this is probably workable. But for the fastest-moving and most technically specialised use cases - in our case, AI cybersecurity and AI-assisted software development being the clearest examples - the current framing creates a structural problem.

Licensed energy providers cannot be expected to have equivalent depth in novel AI security architectures, software generation risks, or state-of-the-art defensive tooling. These are not areas in which the energy sector has accumulated decades of expertise. They are areas where the deepest expertise sits outside it entirely - in specialist AI security firms, in software engineering organisations, in cybersecurity experts, and in the international standards community. Licensees can speak with authority about energy operations; they cannot speak with equal authority about why one AI architectural pattern is structurally more secure than another, or what the implications of a given toolchain choice are for a CNI operator's attack surface.

Under the current model, that expertise enters the sandbox only as filtered through the lead applicant's judgement. For technically complex use cases, this is likely to produce findings that

are shallower than they should be and governance processes that are less technically informed than the subject matter requires.

Proposal: A Named Technology Partner Tier

We propose that Ofgem introduce a Named Technology Partner Tier, distinct from the general partnerships category described in Section 3.2. Named Technology Partners would apply directly to contribute to sandbox governance for specific, technically specialised use cases, without requiring a licensee intermediary. We are not proposing that this tier carries the same regulatory accountability as industry applicants - the lead applicant model for that accountability is sound. We are proposing that technical expertise be given a formal path to influence governance that does not require it to be filtered through an entity that may not be positioned to evaluate it. The design of such a tier - its specific rights, responsibilities, and accountability mechanisms - is properly a matter for Ofgem to determine.

This structure also has a practical benefit for licensees themselves. Under the current model, lead applicants bear responsibility not only for their domain expertise - energy operations, customer obligations, regulatory compliance - but also for translating and mediating highly technical AI security input that they are not best placed to assess. A Named Technology Partner Tier removes that burden: licensees can focus on what they know best, Technology Partners can contribute at the level of depth the subject matter demands, and neither is asked to operate outside their competence.

The governance implication of this is equally important. Rather than forcing all expertise through a single channel - with the inevitable flattening that entails - the sandbox effectively operates two parallel tracks: an industry track, led by licensees who understand the energy system, and a technology track, led by Named Partners who understand the AI and security landscape. Ofgem retains oversight of both. The result is broader and more accurate coverage of the problem space, without requiring either party to pretend to expertise they do not have, and without materially increasing the complexity of the governance structure itself. We would suggest that this tier explicitly include organisations with deep CAF and NIS compliance expertise, given the extent to which the proposed use case intersects with existing regulatory obligations under those frameworks.

Precedent: The FCA Sandbox Model Supports This

This is not an untested idea. The FCA's regulatory sandbox has evolved over successive cohorts to accommodate participants who are not FCA-authorized firms but who bring expertise that the sandbox needs. The FCA recognised that restricting meaningful participation to regulated entities limited the quality of the sandbox's technical findings in areas where the deepest expertise sat outside the regulated population. The FCA's experience provides a model for how a Named Partner Tier can operate within a robust regulatory governance structure.

Q2: Use Case Selection Criteria

Are the proposed use case selection criteria (including commercial neutrality, innovation, sector impact, regulatory uncertainty, testability, governance, and data access) appropriate and sufficient to ensure a fair and transparent process? Are there other criteria, safeguards, or considerations we should include?

The Core Criteria Are Sound

The core criteria are appropriate and we support them.

Proposed Addition: Auditability

Auditability should be a named criterion, not an implicit assumption embedded within testability. In a regulated CNI environment, the ability to produce a clear account of what an AI system did and with what effect is not optional - it is a prerequisite for compliance, incident response, and regulatory accountability. A system that performs well on measurable success metrics but leaves no auditable record of how it reached its outputs is not suitable for deployment in an energy sector context, regardless of its trial performance. Note that we do not mean AI "explainability" here - that is a much harder and largely unsolved problem. We mean something simpler and more achievable: a traceable record of where the AI made which changes, sufficient to verify, audit, and continuously improve.

A Note on "Shift Left"

An established principle in secure software development - "shift left" - holds that vulnerabilities are cheaper and safer to address the earlier they are caught. In CNI environments this is especially acute: a flaw that reaches production in operational technology may not be patchable without service disruption. This holds for any use case, not just software. Use cases that move risk detection earlier in the process should be weighted accordingly.

Access to Operational Data

Section 3.5 of the consultation refers to operational data and infrastructure as a rationale for requiring lead applicants to be licensees or OES. We support this rationale but would ask that it be extended, particularly for security.

For the proposed CNI software security use case, the relevant operational data is not only grid operations data or customer data - it is the software development methodologies, toolchains, and DevSecOps practices used within the energy sector. Without access to realistic representations of how energy sector software is actually built - the tools engineers use, the workflows they follow, the compliance processes they operate within - it is not possible to

meaningfully test or evaluate AI-assisted development approaches in that context. The consultation should make clear that this category of operational information is within scope.

Our Proposed Use Case Evaluated Against the Criteria

- **Commercial neutrality:** Published guidelines and standards benefit the whole sector equally
 - **Innovation:** AI-assisted software development security in CNI environments is a novel area; no existing Ofgem, cross-government, or international framework addresses it
 - **Sector impact:** The software attack surface created by AI coding tools extends across every energy operator that uses them; success materially improves the defensive posture of UK energy CNI as a whole
 - **Regulatory uncertainty:** No current framework explicitly addresses the obligations of OES when using AI tools to develop or maintain the software systems that underpin their NIS compliance
 - **Testability:** Clear and measurable success metrics exist - vulnerability detection rates in AI-generated code, audit trail completeness, time-to-detection of unsafe AI behaviours, and reduction in the attack surface exposed by development toolchains
 - **Governance:** The use case explicitly requires cross-industry participation; energy operators provide the operational context, cybersecurity and AI specialists provide the technical depth
-

Q3: Alignment With Other Initiatives

Are the proposed initiatives and the AI Technical Sandbox distinct and complementary? Are there other existing initiatives we should be aware of?

The Multi-Sandbox Architecture Is Well-Designed

The complementary structure Ofgem has outlined - across the AI Technical Sandbox, the AI Reg Lab, the Energy Regulation Sandbox, and the Future Regulation Sandbox - is sensible, and we support it.

International Standards: An Active Frontier, Not a Settled Body of Work

The consultation references only domestic and EU initiatives. This is understandable but creates a meaningful gap, and the nature of that gap is important: as of early 2026, no body - domestic or international - has published authoritative guidance specifically on how to develop software securely using agentic AI tools. This is not an area where Ofgem can adopt existing international standards and apply them to the energy sector. It is an active frontier where the

standards are being written now. The Ofgem sandbox has an opportunity to be a contributor to that process, not a late recipient of it.

The most substantive international work currently underway sits across several bodies. NIST published RFI 2026-00206 in January 2026, explicitly addressing the security considerations for AI agents - the most direct treatment of this question from any major standards body to date. The NIST NCCoE DevSecOps consortium, which includes Google, Microsoft, IBM, and GitLab among its participants, is developing practical guidance on secure software development practices in the context of AI tooling. NIST's Center for AI Standards and Innovation (CAISI) is tracking the broader AI security standards landscape internationally. And DSIT, through its AI Safety Institute and AI Growth Lab, is developing the UK's domestic position on AI governance - necessarily at a level of generality that is insufficient for CNI on its own.

Sagittal submitted a [formal response](#) to [NIST RFI 2026-00206](#), directly addressing the architectural properties that make AI agents safe or unsafe in regulated environments, and we participate in the NCCoE DevSecOps consortium under a CRADA. Sagittal is well-placed to serve as a bridge between Ofgem's sandbox and these international standards processes - ensuring that UK energy sector findings reach the bodies that are actively shaping the standards, and that the sandbox benefits from the most current international thinking in return. This bridging role does not need to wait for the sandbox to launch; it can begin during the design phase.

Ofgem will also wish to consider how the sandbox aligns with NCSC's published guidance on AI security and secure development practices, and with the Cyber Assessment Framework that already governs OES obligations under NIS. Ofgem is better placed than we are to navigate that alignment directly; we note it here because the sandbox's findings will be significantly stronger if they are legible within the existing UK regulatory and assurance landscape from the outset, rather than retrofitted to it afterwards.

Ofgem Should Shape Standards, Not Inherit Them

The same principle applies across all the bodies working in this space - DSIT, NCSC, NIST, and CAISI alike. Each of them is developing frameworks at a level of generality that is, by design, calibrated to the median case. The median case is not CNI. An AI coding governance framework built to work across retail, professional services, and public administration will not, by default, address the consequence profile, regulatory context, or threat model of energy infrastructure software. That gap will not close itself.

Ofgem is uniquely positioned to close this gap. It is the competent authority for the sector most exposed to this risk, with the regulatory standing to set sector-specific standards and the operational access - through its licensees - to ground those standards in real deployment conditions. No economy-wide body has both, and the UK's unified national regulatory structure

for energy gives Ofgem a coherence of authority that is difficult to replicate elsewhere - including in jurisdictions with more fragmented oversight.

The risk is not that any more general standards body's work is inadequate or misguided. The risk is that without deliberate action by Ofgem, CNI software security becomes a downstream footnote in frameworks designed for the median case - adopted after the fact, retrofitted to a context they were not designed for. As a result, they may not yield the protection that CNI deserves.

The sandbox should be explicitly designed to produce a CNI-specific standard on AI-assisted software security - one that Ofgem then actively contributes upstream into DSIT, NCSC, NIST, and CAISI processes. Ofgem should be shaping what those frameworks say about CNI, not inheriting what they say about everything else and working out how to apply it.

Energy Is Best Placed to Lead - and Need Not Do So Alone

Nothing in this response should be read as suggesting that Ofgem or its licensees can or should address CNI software security in isolation. The deepest expertise in AI security architecture, software supply chain risk, and secure development tooling sits largely outside the licensed energy sector, and the sandbox must draw on it deliberately.

The NCCoE consortium model demonstrates how this can work in practice: a convening body sets the agenda and owns the outputs, while participants from industry, standards bodies, and specialist technology firms contribute technical depth. The governance structure we proposed in Q1 - with a Named Technology Partner Tier - is designed precisely to enable this.

Q4: Engagement and Governance

Do you agree with the proposed approach to engaging participants and governing the sandbox? Are there improvements you would suggest?

The Proposed Structure Is Adequate But Calibrated for Conventional Risk

The Steering Group structure, working group model, and open forum consultation mechanism provide a sound governance foundation for conventional regulatory sandbox use cases. We support them.

Our concern is that the structure as proposed is calibrated for thoroughness rather than pace. This is a good plan in almost all cases, but adds too much risk in the AI coding space. AI coding tools are released, updated, and adopted by working engineers - including across the energy sector - continuously and at speed. Both the offensive capabilities AI gives attackers and the

unintentional exposure created by unregulated development tooling evolve faster than annual review cycles can track. A governance structure operating on that cadence will be authoritative on last year's AI capabilities while practitioners are already working with this year's tools. This is not a failure of intent; it is a structural mismatch due to a white-hot technology, and it applies to the sandbox's findings cadence as much as to its governance intake.

Recommendations

Include technical experts as named Steering Group consultants. The Steering Group will be well-qualified to assess regulatory, commercial, and sector impact questions. For technically specialised use cases it should also have access to named technical advisers with defined rights to provide input on architectural and security matters. This is distinct from the Named Technology Partner Tier proposed in Q1; it operates at the governance level rather than the participation level.

Establish a rapid-response mechanism for emerging AI capabilities. The Steering Group should have a standing pathway to assess and intake materially new AI developments on a compressed timeline, without requiring a full review cycle. A defined sub-group with delegated authority to make an initial assessment and recommend short-term action would be sufficient. To illustrate why this matters: the normalisation of safety bypass flags in agentic coding tools - where developers are routinely prompted to disable safeguards entirely - has occurred within the last *three months*. An annual governance cycle would not have caught it in time to inform practice.

Publish findings iteratively, not as a single end-of-cycle report. Shorter review loops with interim publications allow participants and the wider sector to act on findings as they emerge. If this is operationally onerous across the full sandbox, it should at minimum apply to cybersecurity and AI software development use cases, where the pace of change makes annual publication cadence structurally inadequate.

Q5: Timelines and Next Steps

Do you agree with the proposed timeline for the sandbox? Are there aspects of the timeline or sequencing that you would change?

The Pre-Launch Interval Is the Immediate Risk

The proposed sequence - consultation close, Spring decision, Autumn 2026 launch - is administratively reasonable. We do not object to it.

The risk is what happens in the interval. AI coding tools are being adopted by energy sector engineers today, in the absence of any sector-specific guidance on how to use them safely. Toolchain choices, workflow designs, and development practices are being set without regulatory input, and potentially insecure code is being written into CNI systems. By the time guidance is available, the practices it is meant to shape may already be embedded.

Two adjustments would reduce this exposure. First, Ofgem should pre-publish high-level eligibility and use case selection criteria ahead of the full decision document, at least in the areas of cybersecurity and CNI-specific AI use. This allows prospective participants - licensees and Technology Partners alike - to begin scoping work and assembling evidence, materially reducing the time between application open and first findings. Second, in the interim, Ofgem should consider issuing high-level DevSecOps guidance - a baseline of recommended practices and a short list of practices to avoid. Named Technology Partners, as proposed in Q1, are the natural source of input for this work.

Q6: Ethics and Responsible AI

Do you agree with the proposed approach to ensuring ethical and responsible AI use within the sandbox? What additional safeguards or principles should be included?

The Consultation's Ethics Framework Is Sound, But Incomplete

The ethical values cited in the consultation - fairness, transparency, accountability, consumer trust - are the right values for a regulated sandbox to operate by. The commitment to pre- and post-trial ethics reviews, and to ensuring that AI use within trials is responsible and well-governed, is appropriate and we support it.

The Consultation's Ethics Framework Faces Inward, Not Outward

Those values, as currently framed, are properties of how the sandbox itself operates and how AI is used within trials. It leaves a significant gap, however: there is no discussion of what obligations CNI operators carry to the public regarding how they use AI to develop the software that underpins critical national infrastructure.

Transparency, as currently framed, means the sandbox process is open and well-governed. It does not mean that energy companies are required to disclose how AI is being used in their software development pipelines, or what steps they are taking to ensure that AI-generated code in CNI systems is secure. These are different things.

Public Disclosure on Agentic AI Is a National Security Obligation

The software that controls UK energy infrastructure is a strategic asset. When AI tools are used to develop or modify that software, the public and the government bodies responsible for national security have a legitimate interest in knowing that this is happening and that appropriate mitigations are in place. The public interest here is analogous not to a product disclosure requirement, but to the transparency obligations that apply in other domains where private actors manage assets of national strategic importance. Energy utilities are largely local monopolies - market pressure will not drive this; it must be regulated.

We recommend that the sandbox require participating CNI operators to develop public disclosure frameworks as a defined output: what AI tools are being used in software development, what risks they introduce, and what specific mitigations are in place. Ofgem should use the sandbox to establish what adequate disclosure looks like, and then consider whether it should become a baseline expectation for all OES under its remit.

But a disclosure obligation is only as strong as its evidential basis. Every AI contribution to a CNI codebase - written functions, modified configuration, calls to external services - must be attributable, traceable, and distinguishable from human-authored work. Without this, there is no disclosure a regulator, auditor, or security reviewer could meaningfully verify. Structural auditability is not an internal compliance nicety; it is the precondition for any public accountability about how AI is being used in critical infrastructure software. We recommend that auditability be established as a baseline requirement for AI-assisted software development in CNI.

Oversight Quality, Not Oversight Volume

The consultation assumes, reasonably, that human oversight is a safeguard. It is - but only if it is designed well. The dominant oversight model in standard agentic AI tools is per-action permission prompting, which in practice produces permission fatigue: when engineers approve dozens of low-stakes actions in rapid succession, approval becomes reflexive, and some tools have normalised flags that disable safeguards entirely - one widely-used agentic coding tool ships a flag literally named `--dangerously-skip-permissions` that has become routine in developer workflows - and not one that is appropriate for CNI. More approval events do not produce more meaningful oversight. Review points integrated into existing engineering workflows - code review, version control, merge approval - represent a more robust model, and one that the sandbox's ethics framework should explicitly prefer.

Q7: Stakeholder Support

What support would be most useful to participants during and after the sandbox? How should Ofgem use sandbox findings to support broader sector engagement?

Connect Sandbox Participation to Innovation Funding Pathways

Early-stage technology companies are, in many cases, the organisations with the deepest expertise in the most novel AI security domains, and also the least equipped to absorb the cost of extended regulatory engagement. The Ofgem sandbox risks drawing in only established technology vendors - particularly those from the US, who carry existing regulatory relationships - if it does not take comparable steps to lower barriers for UK-based specialist innovators. We recommend that Ofgem formally connect sandbox participation to UK Innovate financing and other similar innovation funding pathways: regulatory endorsement substantially strengthens funding applications from early-stage companies and is a low-cost lever for Ofgem to expand the quality and diversity of expertise available to the sandbox.

Commission a CNI Software Security Reference Framework

Ofgem should commission a practical, sector-specific reference framework for safe AI-assisted software development in energy CNI as a defined sandbox output. It would address the software development lifecycle specifically - the tools engineers use, the workflows they follow, the compliance checkpoints that already exist within energy sector DevSecOps practice - and it would be immediately applicable by any energy operator using AI coding tools. The NCCoE DevSecOps consortium's work and NIST's CAISI RFI provide a natural starting point; the sandbox's role is to adapt and validate that work for the UK energy sector context.

AI Assurance Evidence: The Sandbox Can Define What Good Looks Like

Ofgem expects that firms demonstrate their AI governance posture through structured evidence. The sandbox is the right place to develop what that evidence looks like for AI-assisted software development specifically - producing a standard template that CNI operators can use to demonstrate that their AI coding practices are safe, auditable, and compliant with NIS obligations, developed through participation rather than imposed top-down.

Q8: General Feedback and Closing

Is there any other feedback you wish to provide on the AI Technical Sandbox proposal?

The Proposed Use Case Delivers Against All of Ofgem's Stated Outcomes

Section 3.8 of the consultation sets out four expected outcomes for the sandbox. The CNI software security use case we have proposed maps directly to each of them.

Sector learning. Practical, evidence-based guidelines for AI-assisted secure software development in CNI, applicable immediately across the energy sector, where no analogous sector-specific guidelines currently exist.

Regulatory validation. A direct test of whether existing Ofgem and cross-government frameworks adequately cover the risks created by AI coding tools in CNI environments - and the mechanism to fill the gap where they do not.

Defining the art of the possible. Evidence, from production deployments in regulated industries, that AI systems with structural auditability and process constraints are deployable and practically superior in CNI environments.

Identifying R&D needs and guidance gaps. A rigorous account of where current standards - including NIST SSDF, the NIS Regulations 2018, and Ofgem's own guidance - fail to address the software attack surface and urgency for CNI that AI coding tools have created.

Delaying Oversight on AI Software Development Is Not a Conservative Choice

There is a framing risk in how CNI software security might be perceived within the sandbox design process. It may appear to be a niche technical matter, better addressed by specialists than by a regulatory sandbox, such as DSIT, NCSC, or another economy-wide body's remit. And the political cost of not including it may appear lower than the cost of the design complexity it introduces.

None of these appearances are accurate. The software attack surface created by AI coding tools across UK energy CNI is new, poorly mapped, and expanding with every month that passes without guidance. Nation-state actors are actively exploiting software vulnerabilities in energy infrastructure at scale. The regulatory framework that governs how energy operators manage their software security - the NIS Regulations, Ofgem's competent authority role, the CAF - does not yet address what it means to discharge those obligations when AI tools are generating the code. That gap is not theoretical. It is current and material.

If this sandbox does not address CNI software security, there is no other body currently positioned to do so with both the sector-specific depth and the regulatory authority that the problem requires. Ofgem is the competent authority for the operators most at risk. The sandbox is the tool Ofgem is creating for exactly this kind of problem. Declining to use it here is not the cautious choice; it is the one that leaves a significant and fast-growing gap in UK CNI's defensive posture unaddressed.

Closing

Sagittal AI welcomes the AI Technical Sandbox as a significant and genuinely necessary initiative. The energy sector's exposure to AI-enabled threats is real and growing, and Ofgem is right to act. We are enthusiastic about its ambitions.

We have submitted this response in the hope that it contributes to a sandbox that is technically credible, ambitious in scope, and practically useful to the operators it serves. We stand ready to engage further - in working groups, governance structures, or direct dialogue with Ofgem's team.

We are grateful for the opportunity to contribute.