



**Response to Ofgem: AI Technical Sandbox Consultation Provided by
WeBuild-AI**

Introduction	3
Q1: Eligibility and Participation	4
Q2: Use Case Selection Criteria	5
2.1 Definition of Innovation	5
2.2 Data Readiness as an Explicit Criterion	5
2.3 Transferability and Scalability	5
2.4 AI Risk Profile	5
2.5 Commercial Neutrality in Practice	6
Q3: Alignment with Other Initiatives	7
3.1 Routing Logic Between Initiatives	7
3.2 Learning from the FCA Supercharged Sandbox	7
3.3 DSIT AI Growth Lab	8
3.4 Additional Interfaces	8
Q4: Engagement and Governance	9
4.1 Technical Expertise on the Steering Group	9
4.2 Conflict of Interest Provisions	9
4.3 Formal Input from Working Groups	9
4.4 Incident Response	9
4.5 Partner Representation	9
Q5: Timelines and Next Steps	10
Q6: Ethics and Responsible AI	11
Q7: Stakeholder Support	12
7.1 Early Publication of Template Documents	12
7.2 Pre-Application Support	12
7.3 Data Access Framework	12
7.4 Partner Register	12
7.5 Capacity Building	12
Q8: General Feedback	13
8.1 Sandbox Infrastructure: Reusing Proven Patterns	13
8.2 An Energy-Sector Synthetic Data Catalogue	13
8.3 A Use Case Model and Training Data Registry	14
8.4 Cloud and Data Sovereignty	15
8.5 AI Supply Chain Transparency	15
8.6 Exit and Transition Planning	15
8.7 Accelerated Deployment Pathway for High-Impact Use Cases	15
8.8 International Regulatory Context	17
Concluding Remarks	17

Introduction

WeBuild-AI welcomes the opportunity to respond to Ofgem's consultation on the proposed AI Technical Sandbox. We commend Ofgem for taking a proactive and structured approach to enabling responsible AI adoption in the energy sector, and for the quality and clarity of the consultation document.

WeBuild-AI is a specialist technology consultancy that builds custom, compliant and production ready AI solutions for enterprise organisations. Our work encompasses the architecture of secure, sovereign data & AI infrastructure, the deployment of machine learning & AI systems at scale, the creation of agentic workflows and the development of digital experiences that solve the biggest problems for our customers. We support clients in navigating the technical and regulatory dimensions of AI adoption, whilst ensuring the systems we build are scalable, resilient and safe.

Our leadership team has deep experience of working across the energy system with organisations such as National Grid Electricity Transmission, National Grid Ventures, National Grid Electricity Distribution, Cadent Gas, UK Power Networks, EDF and

We respond to this consultation from the perspective of a technology provider and potential partner within the proposed sandbox framework. Our observations are intended to be constructive, and we hope they assist Ofgem in refining the pilot design to maximise its value for the energy sector and for consumers.

A recurring theme in our response is the importance of learning from the established infrastructure patterns that have already been proven in other UK regulatory sandbox programmes. The Financial Conduct Authority's Supercharged Sandbox, delivered in partnership with NVIDIA and AWS offers a mature and directly transferable model for how production-grade AI testing infrastructure can be provisioned within a regulatory setting. We believe Ofgem has an opportunity to reuse and adapt these proven patterns rather than designing from first principles, significantly reducing delivery risk and accelerating the pilot's time to value, which is imperative given your tight delivery milestones in H2 of 2026.

We recommend five priority enhancements to strengthen the pilot design:

1. Clarify the role, rights and responsibilities of partners within the sandbox
2. Adopt a centrally provisioned infrastructure model to enable consistent and rigorous testing
3. Introduce data readiness as a standalone use case selection criterion
4. Define a minimum ethical and safety evaluation framework
5. Establish a clear pathway from sandbox testing to live deployment

These recommendations are intended to improve the quality of testing, reduce delivery risk, and maximise sector-wide learning.

Q1: Eligibility and Participation

We broadly agree with the proposed eligibility criteria. The restriction of lead applicant status to licensees, market participants, and operators of essential services is a sound basis for ensuring regulatory accountability and maintaining a clear chain of responsibility throughout the pilot.

However, we would respectfully encourage Ofgem to provide greater clarity on the role, rights, and responsibilities of partners within the sandbox framework. The consultation acknowledges the value of partnerships with technology providers, academia, and other innovators, but the current drafting leaves the practical substance of the partner role somewhat undefined. In our experience, the following areas would benefit from further specification:

- The extent to which partners may contribute to trial design, methodology, and evaluation criteria, rather than serving solely as suppliers of technology or resources.
- Whether partners will have direct access to sandbox environments and representative data under appropriate legal agreements, or whether all data access must be mediated through the lead applicant.
- Whether partners will have the ability to present findings or observations to the Steering Group directly, or whether all communication must be channelled through the lead participant.

We raise these points because technology providers and AI specialists frequently hold the deepest expertise in AI system design, testing methodology, safety evaluation, and risk assessment. If the partner role is structurally subordinate with limited visibility or influence, there is a risk that trials are less technically rigorous than they might otherwise be.

We would also suggest that Ofgem consider publishing minimum expectations for partner involvement in applications. Without such guidance, there is a risk that some lead applicants may treat the partnership element as a conventional procurement exercise, selecting partners primarily on cost rather than on the depth and quality of their technical contribution.

Finally, the principle of commercial neutrality is important and well-founded. We would note, however, that it operates in both directions. Partners contributing significant intellectual property, proprietary methodologies, or specialist expertise will need assurance that published findings do not inadvertently expose their methods or approaches to competitors. We would welcome further detail on the intellectual property protections available to partners, as distinct from those afforded to lead participants.

Q2: Use Case Selection Criteria

The six proposed criteria represent a reasonable and well-considered starting framework. We offer the following observations in the spirit of strengthening the selection process.

2.1 Definition of Innovation

The current formulation requires that a proposal should address something “not currently possible in the live market.” Whilst we understand the intent, this sets a high threshold that may inadvertently exclude valuable use cases where the underlying AI approach is technically mature but cannot be deployed at scale owing to regulatory uncertainty, data access constraints, or safety considerations. We would suggest reframing this criterion to encompass proposals that are “not currently deployed or deployable at scale in the live market owing to regulatory, technical, or safety uncertainty.” This broadening would ensure the sandbox captures genuinely important use cases that are blocked by non-technical barriers.

Indeed, established use cases could still be optimised and delivered with more performant, capable and cost efficient models, with the ability to be run on different devices. For instance, mobile, tablet, on edge devices and subsequently require less compute. It would therefore be wise not to rule out established use cases, in this instance.

2.2 Data Readiness as an Explicit Criterion

The consultation references data access within the testability criterion, but we would suggest that data readiness merits recognition as a standalone consideration. Many promising use cases will be difficult to progress if representative operational data cannot be obtained, is subject to existing sharing restrictions, or requires extensive anonymisation. Requiring applicants to evidence a clear and credible data access plan at the point of application would help Ofgem identify and address data barriers early in the process.

We address the question of how the sandbox might proactively facilitate data access, including through synthetic data provision, in our response to Q8 below.

2.3 Transferability and Scalability

We would encourage Ofgem to consider adding a criterion relating to the transferability or scalability of proposed use cases. The pilot will deliver the greatest return if its learnings can be generalised across multiple licensees, market segments, or system functions, rather than remaining specific to a single organisation’s circumstances. Favouring use cases with broad applicability would maximise the value of the public investment in the sandbox.

2.4 AI Risk Profile

The current criteria do not explicitly require applicants to articulate the risk profile of their proposed AI system. We would suggest that applications should include an assessment of the system’s autonomy level, the potential consequences of failure, the sensitivity of the data involved, and the degree of human oversight envisaged. This information would enable Ofgem and the Steering Group to apply proportionate oversight. A demand forecasting model and an autonomous grid-balancing agent, for example, present materially different risk profiles and would warrant different levels of scrutiny.

Furthermore, learnings from financial services can be taken from domains such as operational resilience for important business services and the planning & testing for extreme yet plausible scenarios. We propose that Ofgem considers how it can use agents & large language models to review use cases and support in the drafting of extreme yet plausible disruption scenarios as a consequence of the proposed AI system behaving erroneously and outside of its constraints.

This type of solution should be lightweight, but rigorous enough to help devise testing strategies that can be evolved during the AI sandbox development process so that risks can be validated and better understood before further investment.

These findings should be made accessible as part of lessons learned knowledge graph that maps use cases to their testing results combining agents and LLM's to support future testing and validation of evolving use cases that are considered for AI sandbox assessment.

2.5 Commercial Neutrality in Practice

The commercial neutrality criterion is rightly included. We would note, however, that complete neutrality may be difficult to achieve in practice where a trial produces results that clearly benefit the lead participant. The more important question is perhaps whether the learnings from each trial are shared widely and specifically enough that other market participants could replicate or build upon them. We would welcome further guidance from Ofgem on how this balance will be managed.

Q3: Alignment with Other Initiatives

The consultation provides a helpful overview of the various initiatives with which the AI Technical Sandbox is intended to align. We offer the following observations on how this alignment might be strengthened.

3.1 Routing Logic Between Initiatives

The document references the AI Reg Lab, Energy Regulation Sandbox, Future Regulation Sandbox, NESO programmes, UKRI and SIF/NIA portfolios, the FCA's sandbox experience, and DSIT's AI Growth Lab. Whilst each is acknowledged, the consultation does not clearly articulate the decision criteria or triggers for routing evidence, findings, or participants between them. We would suggest that Ofgem publish a clear routing framework, ideally in diagrammatic form, setting out the conditions under which findings from the AI Technical Sandbox would be escalated to or shared with each of the adjacent initiatives, who would be responsible for those decisions, and what timelines would apply.

3.2 Learning from the FCA Supercharged Sandbox

The consultation rightly references the FCA as a leader in regulatory sandboxes. We would encourage Ofgem to go further than drawing on the FCA's governance model, and to examine the practical infrastructure that underpinned the FCA's Supercharged Sandbox in detail. The Supercharged Sandbox, delivered through a partnership between the FCA, NVIDIA, and AWS, offered participating firms access to secure cloud environments with GPU-accelerated compute, enriched synthetic datasets, and production-grade developer tooling from the outset of the testing programme.

This infrastructure-first approach proved transformative. The FCA's first Supercharged Sandbox cohort selected 23 firms from 132 applications and ran a structured 20-week testing programme. Because teams had early access to production-grade compute and data whilst their systems were still in development, they were able to surface issues that are typically deferred in conventional pilot programmes, including edge-case behaviour, explainability trade-offs, data limitations, and operational readiness. Several participating firms reported that this approach led them to rethink architectures, narrow scope, or de-risk agentic approaches in ways that would not have occurred without sustained, rigorous testing in a well-provisioned environment.

We would strongly encourage Ofgem to explore whether the infrastructure patterns established by the FCA, NVIDIA and AWS can be reused or adapted for the energy sector sandbox. The core components of that infrastructure, namely secure multi-tenant cloud environments, GPU-accelerated compute for model training and evaluation, curated synthetic datasets, and streamlined onboarding tooling, are largely sector-agnostic and could be extended to accommodate energy-specific data types and use cases. Reusing established patterns would materially reduce the delivery risk and procurement timeline for the pilot, and would avoid the cost and delay of designing bespoke infrastructure from first principles.

We recognise that the energy sector presents different risk characteristics from financial services, including physical safety considerations, infrastructure resilience requirements, and the real-time nature of system operation. These differences will require sector-specific

adaptations, particularly in the areas of data types, simulation environments, and safety evaluation criteria. However, the underlying infrastructure platform and operating model from the FCA programme provide a sound and proven foundation upon which those adaptations can be built.

3.3 DSIT AI Growth Lab

We note that the consultation explicitly defers consideration of alignment with DSIT's AI Growth Lab, stating that it is not the focus of the current consultation. We respectfully suggest that this interface warrants closer attention. The AI Growth Lab is likely to set cross-economy standards and expectations for AI sandboxing. If the energy-sector sandbox is designed without reference to those emerging standards, there is a risk of incompatibility or duplication. At minimum, we would encourage Ofgem to commit to interoperability with the Growth Lab, including the adoption of common taxonomies, shared evaluation frameworks, and mutual recognition of sandbox findings where appropriate.

3.4 Additional Interfaces

We note that the consultation does not reference alignment with the Information Commissioner's Office on data protection aspects of AI testing, or with the Department for Energy Security and Net Zero on policy alignment. Both would appear to be important interfaces for the sandbox and we would welcome their inclusion in the final framework.

Q4: Engagement and Governance

The proposed governance structure is sensible in its overall design. We offer the following suggestions for strengthening it.

4.1 Technical Expertise on the Steering Group

We would strongly encourage Ofgem to ensure that the Steering Group includes members with deep technical expertise in artificial intelligence, not solely energy sector or regulatory representation. Without dedicated AI expertise in the governance structure, there is a risk that use case selection and evaluation are driven primarily by policy or commercial considerations rather than technical merit and safety. This is particularly important given the rapid pace of development in AI capabilities and the corresponding evolution of risk profiles.

We recommend at least one independent AI safety or assurance expert sits on the Steering Group as well.

4.2 Conflict of Interest Provisions

The consultation references conflict of interest provisions but does not specify them in detail. Given that the Steering Group will oversee the selection of use cases in which its members' organisations may have a direct or indirect interest, we would suggest that Ofgem publish the Steering Group's Terms of Reference alongside the decision document, rather than at a later stage. This would provide stakeholders with confidence that the governance arrangements are sufficiently robust before applications open.

4.3 Formal Input from Working Groups

Working groups and open forums are a welcome feature of the proposed engagement model. We would suggest that these groups be given a formal mechanism to escalate concerns or recommendations to the Steering Group, to ensure that wider stakeholder input has a meaningful pathway into decision-making rather than serving a purely consultative function.

4.4 Incident Response

The governance structure is currently silent on what happens in the event that a trial produces unintended harms, a data incident occurs, or findings indicate that a participant's AI system is unsafe. We would recommend that Ofgem develop and publish a clear incident response and escalation protocol before the pilot commences. This would provide assurance to all parties and would support the orderly management of any difficulties that arise during testing.

4.5 Partner Representation

We would suggest that technology partners be afforded some form of representation or structured input channel to the Steering Group, even if they are not appointed as full members. Partners will often hold the closest understanding of the technical characteristics and limitations of the systems under test, and their perspective would be valuable in informing governance decisions.

Q5: Timelines and Next Steps

The proposed timeline is ambitious but, in our view, achievable with careful planning. We offer the following observations.

The period between consultation close in March 2026 and the anticipated pilot launch in Autumn 2026 will need to accommodate the publication of a decision document, the development and issuance of participation guidance, the application and selection process, the negotiation of legal agreements, and the establishment of sandbox environments. This is a substantial programme of work within approximately six months, and we would encourage Ofgem to publish an indicative milestone schedule with clearly identified dependencies in the decision document. This would enable prospective participants and partners to plan their resource allocation with greater confidence.

We would reiterate that the timeline for infrastructure provisioning could be significantly compressed if Ofgem were to reuse the established cloud infrastructure and operating model from the FCA Supercharged Sandbox programme, rather than procuring bespoke infrastructure.

We would also suggest that Ofgem clarify whether the 12-month pilot duration runs from the date on which the sandbox environment becomes available, or from the commencement of the first trial. If trials begin at different times, which is likely given a sequential application process, the effective duration of the pilot may extend considerably beyond 12 months.

A phased intake model may be worth considering. An initial cohort of two or three use cases would allow Ofgem to stress-test the governance framework and technical infrastructure before admitting a second cohort. This approach would reduce the risk associated with launching multiple trials simultaneously and would allow early lessons to be incorporated into the operation of subsequent trials.

Finally, we would welcome the inclusion of a formal interim evaluation point at approximately the six-month mark, with published interim findings. This would provide an opportunity to assess progress, make adjustments, and maintain stakeholder engagement throughout the pilot.

Q6: Ethics and Responsible AI

The consultation's emphasis on ethical considerations is welcome and appropriate. We offer the following suggestions for strengthening the ethical framework underpinning the pilot.

The consultation references pre- and post-test ethics and safety reviews, but does not specify what these reviews will cover, who will conduct them, or what standards they will apply. Without this detail, there is a risk of inconsistent or superficial ethical review across trials.

We would encourage Ofgem to define or adopt a minimum ethical evaluation framework for the sandbox, covering at least the following areas:

- Bias and fairness testing across protected characteristics.
- Explainability requirements, proportionate to the risk profile of the AI system under test. This may need to be mapped to the EU AI Act and where possible align to industry best practices such as ISO 42001. Indeed, any ethics and compliance approach needs to be proportionate to ensure that innovation is not stifled.
- Use case analysis needs to be as dynamic as possible and where it is safe to do so, leverage AI to fast track the risk assessment process.
- Data protection impact assessments.
- Consumer impact assessments.
- Security testing, including adversarial robustness where appropriate.
- We recommend Ofgem adopts a minimum evaluation baseline including bias testing, explainability proportionality, DPIAs, and adversarial testing for high-risk systems

The consultation mentions red-teaming in the context of the AI Reg Lab but does not establish it as a requirement for sandbox trials. For higher-risk AI applications, structured adversarial testing steps might need to be a mandatory element of the evaluation process. We would suggest that Ofgem specify minimum red-teaming requirements proportionate to each system's risk profile. This is all the more important at a time when there have been well publicised AI system disruptions as a result of nefarious 3rd party attacks. For instance, the Blue-Hat attack on McKinsey's Lilli system that was reported in March 2026.

We also note that the current framework contemplates ethics reviews only at the pre-test and post-test stages. AI systems can drift or produce unexpected behaviours during operation. The sandbox may want to employ continuous monitoring throughout each trial, with defined alert thresholds and a clear process for pausing or terminating a trial if monitoring identifies concerns.

We would additionally suggest that participants be required to publish an AI transparency statement, equivalent to a model card, for each system tested in the sandbox. This would support the broader objectives of sector-wide learning and would establish a useful precedent for transparency in energy-sector AI deployment. This would also allow us to track use cases and identify synergies between proposed solutions or duplications of effort that may be missed as a result of human assessments in isolation.

Q7: Stakeholder Support

We welcome Ofgem's commitment to supporting stakeholders throughout the pilot and would offer the following practical suggestions.

7.1 Early Publication of Template Documents

We would encourage Ofgem to publish draft legal agreements, application forms, and data-sharing agreement templates as early as possible in the process. Smaller organisations, including SME technology providers, require adequate lead time to review these documents with legal counsel. If complex legal terms are introduced only at the application stage, this may inadvertently disadvantage smaller and more innovative participants.

7.2 Pre-Application Support

A pre-application support process, offering informal and non-binding discussions in which potential applicants and partners can test whether their proposed use case is within scope, would be of considerable value. The FCA offers a similar facility and, in our understanding, it has been effective in improving the quality of applications and reducing wasted effort on both sides.

7.3 Data Access Framework

One of the most significant practical barriers to meaningful sandbox testing will be the availability of representative operational data. We would encourage Ofgem to work with NESO and network licensees to establish pre-agreed data-sharing protocols or to facilitate the provision of synthetic data for common use cases. A clear data access framework would materially improve the quality and pace of trials.

7.4 Partner Register

Ofgem might consider establishing an opt-in partner register: a directory of technology providers, academic groups, and other organisations that have expressed an interest in supporting sandbox trials. This would help licensees and market participants identify suitable partners, and would improve the visibility of smaller innovators who may not have existing relationships with the major energy companies.

7.5 Capacity Building

Many energy-sector organisations are at a relatively early stage in their AI maturity. Ofgem could usefully publish educational materials or host briefing sessions covering AI testing methodology, risk assessment frameworks, and evaluation best practice. Such capacity building would benefit both prospective participants and the wider sector, and would support the sandbox's objective of generating sector-wide learning. These could also be delivered as part of accelerator trials where external skills are provided by reputable partners who embed their experts into the teams of participants to ensure that any initial onboarding steps are hassle free.

Q8: General Feedback

8.1 Sandbox Infrastructure: Reusing Proven Patterns

The consultation does not explicitly address whether Ofgem will provide compute resources, data environments, testing tools, or simulation platforms, or whether participants will be expected to supply their own infrastructure. This is, in our view, the single most important design question for the pilot. The term “sandbox” implies an environment, but the consultation as currently drafted reads more as a governance and oversight framework around self-hosted trials. If the latter is the intended model, it would be helpful for Ofgem to state this explicitly so that organisations can plan accordingly.

We would strongly recommend that Ofgem adopt the former model and provide a shared, centrally provisioned testing environment. The FCA’s experience with the Supercharged Sandbox demonstrates that providing teams with early access to production-grade infrastructure, including GPU-accelerated compute, curated datasets, and developer tooling, fundamentally changes the quality and rigour of testing outcomes. Teams that had access to this infrastructure from the outset reported that they were able to move beyond theoretical feasibility testing and into genuine operational viability assessment, surfacing issues with edge-case behaviour, explainability, and data limitations that would not have been identified in a self-hosted environment.

The infrastructure platform and operating model developed for the FCA, in partnership with NVIDIA and AWS, is already operational and has been tested at scale with a cohort of 23 firms. Its core components, namely secure multi-tenant cloud environments, high-performance compute, synthetic data provision, and streamlined onboarding, are largely sector-agnostic. Rather than procuring or building bespoke infrastructure, Ofgem could explore whether this existing platform can be extended or adapted for energy-sector use cases, significantly reducing procurement timelines, delivery risk, and cost.

8.2 An Energy-Sector Synthetic Data Catalogue

We would like to propose a specific initiative that we believe could substantially enhance the value and accessibility of the AI Technical Sandbox: the creation of an energy-sector synthetic data catalogue.

Data availability is consistently identified as one of the most significant barriers to AI development in regulated industries. The FCA addressed this challenge within its Supercharged Sandbox by providing participating firms with enriched synthetic datasets that allowed meaningful testing without exposing real consumer or market data. The energy sector faces an equivalent challenge, arguably a more acute one, given the diversity of data types involved, spanning smart meter readings, grid telemetry, generation output, market pricing, weather data, and consumer interaction records, among others.

We would suggest that Ofgem work with NESO, network licensees, and academic partners to commission the development of curated, representative synthetic datasets aligned to the most common and highest-value use cases anticipated for the sandbox. These datasets would serve several purposes:

- They would enable prospective applicants to begin prototyping and developing their AI systems before formal acceptance into the sandbox, materially improving the quality and readiness of applications.
- They would provide a common baseline against which different approaches to the same use case could be evaluated, supporting the sandbox's objectives of sector-wide learning and comparability.
- They would reduce the time spent within the sandbox on data preparation, anonymisation, and access negotiations, allowing teams to focus on the substantive work of model development, testing, and evaluation.
- They would lower barriers to entry for smaller participants and partners who may not have independent access to large-scale operational datasets.

The synthetic data catalogue should be designed with extensibility in mind. As new use cases are admitted to the sandbox, corresponding synthetic datasets could be developed and added to the catalogue, building a growing resource for the sector over time. Ofgem might also consider inviting academic institutions to contribute to the development and validation of these datasets, both to improve their quality and to strengthen the sandbox's links with the research community.

8.3 A Use Case Model and Training Data Registry

Building on the synthetic data catalogue, we would propose that Ofgem establish a model and training data registry as a core output of the sandbox programme. The purpose of this registry would be to create a structured, publicly accessible catalogue of the AI models developed and tested within the sandbox, together with the training data, evaluation methodologies, and performance metrics associated with each use case.

This registry would function as follows:

- Upon completion of a sandbox trial, participants would be invited (or, for non-confidential elements, required) to publish a standardised record of their model to the registry. This record would include the model's purpose and scope, the architecture and approach used, the training data employed (or a reference to the relevant synthetic dataset from the catalogue), the evaluation methodology, and the key performance and safety metrics observed during testing.
- Each registry entry would be tagged against the specific use case it addresses, enabling other organisations to discover, review, and learn from previous work on similar challenges. This would directly support the consultation's stated objective of sector-wide learning.
- Where participants consent and intellectual property protections permit, the registry could also host the trained model weights or evaluation artefacts themselves, enabling other teams to reproduce results, benchmark their own approaches against an established baseline, or use a pre-tested model as a starting point for further development.
- The registry would include a clear provenance chain for each entry, linking the model to its training data, its evaluation results, and the sandbox trial in which it was tested. This would support transparency, auditability, and the development of trust in sandbox outputs over time.

The value of such a registry would compound with each successive sandbox cohort. Over time, it would build into a substantial, publicly accessible knowledge base of tested AI approaches for the energy sector, complete with the data, methods, and evidence needed for others to learn from, replicate, or extend previous work. This would be a genuinely distinctive output that goes beyond what any individual organisation could produce in isolation, and it would position the Ofgem sandbox as a sector-leading initiative in terms of practical, reusable outputs.

We recognise that the design of such a registry raises important questions around intellectual property, commercial sensitivity, and data governance, which would need to be addressed carefully in the participation agreements. However, we believe these challenges are manageable, particularly if the registry is designed from the outset with tiered disclosure levels: a mandatory public summary layer, an optional detailed technical layer, and a restricted layer for confidential elements accessible only to Ofgem and the Steering Group.

8.4 Cloud and Data Sovereignty

Given the sensitivity of energy operational data, we would suggest that Ofgem specify requirements around data residency, cloud hosting standards, and security accreditation for sandbox environments. As a baseline, UK-hosted infrastructure with appropriate certification, such as Cyber Essentials Plus and ISO 27001, would be a reasonable expectation. If Ofgem were to adopt the AWS and NVIDIA infrastructure model established by the FCA, these requirements could be met through existing UK-region cloud services with well-understood compliance postures. Clear requirements in this area would provide participants and partners with the certainty needed to design compliant architectures from the outset.

Indeed, by leveraging NVIDIA solutions, there is also the option to explore how AI solutions can be deployed through Neo Cloud providers like N-Scale

8.5 AI Supply Chain Transparency

The consultation does not address the risk that AI systems tested in the sandbox may rely on third-party foundation models or cloud-hosted AI services with opaque data handling practices. We would suggest that Ofgem require participants to disclose their AI supply chain as part of the application process, including model provenance, training data governance arrangements, and material third-party dependencies. This transparency would support informed risk assessment and would be consistent with the emerging direction of AI governance policy in the United Kingdom and internationally.

8.6 Exit and Transition Planning

The consultation is largely silent on the pathway from sandbox to live deployment. What happens at the conclusion of a trial? Under what conditions, if any, may participants deploy a tested system into production? What ongoing obligations, if any, attach to a system that has been evaluated through the sandbox? These are amongst the most important practical questions for prospective participants, and we would encourage Ofgem to address them in the decision document or in subsequent participation guidance.

8.7 Accelerated Deployment Pathway for High-Impact Use Cases

We would encourage Ofgem to consider, as part of the sandbox design, the establishment of a pre-defined accelerated deployment pathway for use cases that demonstrate exceptional and unequivocal value during testing.

The energy sector faces a number of challenges where the timely deployment of a proven AI solution could yield transformative benefits for consumers, the system, and the nation's progress towards net zero. These include, by way of illustration, the optimisation of generation dispatch and demand response to reduce wholesale energy costs and lower household bills; the identification and protection of vulnerable customers through predictive analytics, particularly during periods of price volatility or supply disruption; and the real-time management of grid stability and emissions intensity as the share of intermittent renewable generation continues to grow.

It is conceivable that a use case tested within the sandbox could produce results of such clarity and significance that delaying its deployment through a standard post-pilot review and transition process would represent a material missed opportunity. This is particularly the case in the context of exigent circumstances. The experience of the Covid-19 pandemic, severe weather events, and supply-side shocks has demonstrated that the energy system can be subject to acute, unforeseen stresses in which the rapid availability of effective AI-driven tools could meaningfully improve outcomes for consumers and system operators alike. A black swan event of this nature may create an urgent operational need for a capability that has already been validated within the sandbox but has not yet completed the full transition to production deployment.

We are not suggesting that the rigour of the sandbox evaluation process should be compromised. Rather, we would propose that Ofgem design, in advance, a structured fast-track mechanism that can be invoked when a use case meets clearly defined criteria. Such a mechanism might operate as follows:

- Threshold criteria for accelerated deployment would be agreed and published at the outset of the pilot. These criteria might include: the use case has completed all mandatory safety and ethics evaluations with no outstanding concerns; the performance metrics observed during testing meet or exceed pre-agreed thresholds by a significant margin; the use case addresses a demonstrable and time-sensitive need relating to consumer protection, system resilience, emissions reduction, or energy affordability; and the Steering Group has reviewed the evidence and provided a positive recommendation.
- A condensed transition protocol would be defined, specifying the minimum additional assurance steps required before production deployment. This protocol would be shorter than the standard post-pilot pathway but would retain the core safeguards, including a final safety review, a data protection assessment, and agreement on ongoing monitoring obligations. The aim would be to compress the transition timeline from months to weeks where the evidence supports it.
- Ofgem would retain full discretion over whether to invoke the accelerated pathway in any given case. The mechanism would be permissive rather than mandatory: it would create the option for rapid deployment where the evidence is compelling, without creating any expectation or entitlement on the part of participants.

- Where the accelerated pathway is invoked in response to an emergency or exigent circumstance, the deployment could be authorised on a time-limited basis, subject to enhanced monitoring and a requirement for a full post-deployment review within a defined period. This would enable the benefits of the AI system to be realised during the period of acute need, whilst preserving the opportunity for thorough evaluation under normal operating conditions.

The value of designing this pathway in advance, rather than improvising it under pressure, cannot be overstated. If an emergency arises and no pre-agreed mechanism exists for moving a validated sandbox use case into production at speed, the most likely outcome is either that the opportunity is missed entirely, or that deployment occurs through ad hoc arrangements without the benefit of structured oversight.

Neither outcome serves the interests of consumers or the sector. A pre-defined accelerated pathway, with clear criteria, published safeguards, and retained regulatory discretion, would ensure that the sandbox is not only a vehicle for learning but also a mechanism for delivering tangible value to the energy system when it is most needed.

We would also note that the existence of such a pathway would strengthen the sandbox's proposition to prospective participants. Organisations are more likely to invest the time, resource, and expertise required to develop and test a high-quality AI solution within the sandbox if there is a credible route to deployment for use cases that demonstrate exceptional results. Without such a route, there is a risk that the sandbox is perceived as a purely academic exercise, which would limit both the quality of applications and the practical impact of the pilot.

Finally, at a time when AI is seen to be of utmost national importance, we also need to ensure that the United Kingdom has the ability to move at speed and is seen as a progressive country to build and deploy AI systems within. Indeed, there is always the risk that the details of use cases are capitalised on by other Nations and 3rd parties and become services that our energy system operators must procure and pay for. As opposed to developing and building valuable IP that could be monetised in the future for re-sell to energy operators in other countries or jurisdictions.

8.8 International Regulatory Context

The EU AI Act is now in force and will affect AI systems operating across the GB-EU boundary. Whilst UK law does not currently impose equivalent requirements, many energy-sector participants operate in both markets. We would suggest that Ofgem consider whether sandbox evaluation criteria should map to the EU AI Act's risk categories, at least on an informational basis, to support participants who require compliance across multiple jurisdictions. This would also position the UK energy sector well in the event that domestic AI legislation is introduced.

Concluding Remarks

We regard the AI Technical Sandbox as a positive and needed initiative. The energy sector's adoption of AI is accelerating, and a structured, regulatory-led approach to testing and evaluation is greatly preferable to the alternative of fragmented, ad hoc arrangements.

Our principal recommendation is that Ofgem look closely at the infrastructure, operating model, and data provisioning approach that has already been proven through the FCA's Supercharged Sandbox programme. Reusing these established patterns, adapted for the specific requirements of the energy sector, would reduce delivery risk, compress timelines, and ensure that the sandbox provides a genuinely capable testing environment rather than a governance wrapper around self-hosted trials. Complementing this with an energy-sector synthetic data catalogue and a use case model registry would create a set of durable, publicly accessible assets that compound in value with each successive cohort.

WeBuild-AI would welcome the opportunity to engage further with Ofgem and with prospective lead participants as the pilot progresses. We are committed to supporting responsible AI adoption in the energy sector and believe the sandbox has the potential to make a meaningful contribution to that objective.

Submitted by:



Ben Saunders

Co-Founder

ben.saunders@webuild-ai.com



Mark Simpson

Co-Founder

mark.simpson@webuild-ai.com