How the Proposed Architecture Can Support the EDPB Guidelines on Web Scraping for Generative AI

Executive Summary

The European Data Protection Board (EDPB) Draft Guidelines on the processing of personal data through web scraping for the training of generative AI emphasise that controllers should implement appropriate **technical and organisational measures (TOMs)** throughout the AI lifecycle. The Guidelines repeatedly refer to purpose limitation, data minimisation, transparency, accountability, protection of special categories of personal data, and safeguards against unlawful disclosure.

While the Guidelines identify these objectives, they do not prescribe a protocol-level technical mechanism capable of enforcing them during data processing or AI execution. Instead, they leave the choice of implementation to controllers.

The architecture disclosed in **WO 2026/150382 – The DAS Protocols** proposes a hardware-rooted execution-governance framework that can complement existing compliance measures by introducing protected validation before data use, disclosure, export, computation, or any other externally effective operation. Rather than relying solely on policies, contractual obligations, application permissions, or post-hoc auditing, the architecture enables controller-defined processing conditions to be verified before execution proceeds.

Accordingly, the architecture is not presented as replacing the GDPR or determining legal compliance. Instead, it provides technical infrastructure that may assist controllers in implementing GDPR technical and organisational measures through protected pre-effectuation enforcement.

The Problem Identified by the EDPB

The draft Guidelines recognise several practical challenges associated with web scraping for AI training.

These include:

- collection of personal data without the knowledge of individuals;
- reuse of data beyond the original collection purpose;
- large-scale processing of heterogeneous datasets;
- collection of unnecessary personal information;
- accidental collection of special categories of personal data;

- cross-border processing of information;
- difficulties demonstrating accountability;
- risks of memorisation, regurgitation and unintended disclosure by AI systems.

The Guidelines therefore encourage controllers to implement technical safeguards capable of reducing these risks throughout the AI development lifecycle.

The Missing Technical Layer

Most existing AI governance mechanisms operate at one of four levels:

- organisational policies;
- contractual obligations;
- software access controls;
- post-processing auditing.

These approaches influence behaviour but generally assume that once software has access to data, processing may proceed.

The proposed execution-governance architecture introduces an additional infrastructure layer positioned immediately before externally effective processing.

Instead of treating possession of data as authority to process it, the architecture separates:

- possession of data;
- authority to process;
- authority to disclose;
- authority to export;
- authority to produce externally effective consequences.

Every requested operation first enters a protected non-effective state.

Only after controller-defined validation predicates have been successfully verified within a protected enforcement environment can execution proceed through the controlled execution boundary.

If validation fails, the requested operation remains non-effective and no externally effective consequence occurs.

Supporting Purpose Limitation

The EDPB emphasises that personal data collected for one purpose should not automatically be reused for another incompatible purpose.

Traditionally, purpose limitation is implemented through policies, contracts or organisational governance.

The proposed architecture enables purpose to become a protected execution predicate.

Before an AI system can use personal data, controller-defined validation may verify that:

- the requested processing purpose is authorised;
- the intended processing falls within approved policy;
- the processing environment satisfies applicable requirements.

If these predicates are not satisfied, execution can be denied before processing becomes externally effective.

This transforms purpose limitation from a documentation requirement into an enforceable execution condition.

Supporting Data Minimisation

The Guidelines recommend limiting the collection and processing of personal data to what is necessary.

Current approaches primarily focus on preprocessing datasets using filtering, deletion, anonymisation or synthetic data generation.

The proposed architecture complements these approaches by introducing execution-time verification.

Controller-defined predicates may specify:

- authorised data categories;
- permitted processing operations;
- approved datasets;
- approved recipients;
- authorised AI models.

Only requests satisfying these conditions are permitted to proceed.

This provides an additional enforcement layer beyond preprocessing alone.

Supporting Transparency and Accountability

The EDPB encourages controllers to demonstrate that appropriate safeguards have been implemented.

The proposed architecture supports this objective through protected validation evidence generated before execution.

Such evidence can demonstrate that controller-defined validation predicates were successfully verified before an externally effective operation occurred.

This differs from purely retrospective audit logging because validation occurs before execution rather than after the fact.

Supporting Protection Against Unauthorised Disclosure

The Guidelines discuss safeguards preventing unlawful disclosure of personal information.

Within the proposed architecture, disclosure itself becomes a protected operation.

Before personal information can be disclosed, exported or released, controller-defined validation conditions may be verified.

If those conditions fail, disclosure remains non-effective.

Consequently, disclosure authority is separated from mere possession of the underlying information.

Supporting Cross-Border Processing

Many AI systems involve international movement of personal information.

The proposed architecture allows controller-defined jurisdictional predicates to become execution conditions.

These may include verification of:

- destination jurisdiction;
- approved recipient;
- organisational authority;
- applicable regulatory constraints;

- export restrictions;
- policy version.

Only operations satisfying these requirements proceed through the execution boundary.

Supporting AI Training

Within AI training environments, protected validation predicates may include:

- authorised training purpose;
- approved training dataset;
- approved model identity;
- authorised training environment;
- jurisdiction;
- approved processing policy;
- retention status;
- revocation status;
- organisational authorisation.

Every training request may therefore be validated before becoming operationally effective.

Supporting Protection of Special Categories of Personal Data

The Guidelines recognise that processing special categories of personal data generally requires additional legal justification under Article 9 GDPR.

The proposed architecture does not determine whether processing is lawful.

However, controller-defined validation predicates may require verification that:

- processing of special categories has been authorised;
- appropriate safeguards have been configured;
- required organisational approvals exist;
- applicable policy conditions have been satisfied.

Consequently, processing may be prevented before execution where required validation conditions are absent.

Supporting AI Risk Mitigation

The EDPB also discusses technical measures reducing risks such as memorisation, regurgitation and unintended disclosure.

The proposed architecture does not replace model-level mitigation techniques.

Instead, it introduces an additional infrastructure control layer capable of governing:

- authorised data access;
- authorised computation;
- authorised disclosure;
- authorised export;
- authorised externally effective operations.

This allows execution governance to complement existing AI safety techniques rather than replacing them.

Relationship to Existing GDPR Compliance

The proposed architecture should not be interpreted as establishing GDPR compliance automatically.

Compliance remains dependent upon the controller's legal obligations, governance processes, organisational measures and regulatory assessment.

Instead, the architecture provides technical infrastructure capable of assisting controllers in implementing technical and organisational measures supporting principles including:

- purpose limitation;
- data minimisation;
- accountability;
- controlled disclosure;
- jurisdiction-aware processing;
- execution governance;

- policy enforcement.

Accordingly, the contribution is technical rather than legal.

Industrial Applicability

The execution-governance architecture is applicable across a broad range of AI and digital infrastructure environments, including:

- generative AI training platforms;
- foundation model infrastructure;
- confidential computing environments;
- cloud computing;
- telecommunications;
- cross-border data processing;
- government information systems;
- healthcare AI;
- financial infrastructure;
- critical infrastructure;
- industrial automation;
- autonomous systems.

The architecture is infrastructure-agnostic and may operate alongside existing identity systems, policy engines and AI governance frameworks.

Conclusion

The EDPB Guidelines identify an important need for stronger technical safeguards governing the processing of personal data in AI systems. They emphasise that controllers should implement appropriate technical and organisational measures but intentionally remain technology-neutral regarding how such safeguards should be realised.

The execution-governance architecture disclosed in **WO 2026/150382 – The DAS Protocols** proposes one possible technical implementation. By introducing protected pre-effectuation validation, controller-defined processing conditions can be verified before data use, disclosure, export, computation or other externally effective operations occur.

Rather than replacing existing legal obligations, organisational governance or AI safety techniques, the architecture provides an additional protocol-level enforcement layer capable of strengthening technical implementation of controller-defined safeguards. As AI systems continue to evolve toward increasingly autonomous and distributed operation, infrastructure-level execution governance of this nature may contribute to more robust implementation of GDPR technical and organisational measures while preserving flexibility for controllers to define their own lawful processing policies.

Addressing Legacy-System, Latency and Industry Adoption Considerations

A potential concern regarding protocol-level execution governance is whether introducing protected pre-effectuation validation would be compatible with existing AI infrastructure and whether it would introduce unacceptable latency or operational overhead. These considerations are important for practical deployment and should be distinguished from the underlying technical capability of the proposed architecture.

Legacy Infrastructure

Most existing AI systems were designed around direct data accessibility, distributed storage, and high-throughput processing pipelines. Integrating execution-governance mechanisms into these environments may require modifications to data ingestion workflows, orchestration layers, policy engines, or execution pipelines.

However, the proposed architecture is not intended to require wholesale replacement of existing infrastructure. It may be deployed incrementally alongside existing identity systems, access-control frameworks, confidential-computing environments, or policy engines, allowing organisations to introduce protected execution boundaries progressively where regulatory, commercial, or security requirements justify additional assurance.

Accordingly, legacy infrastructure should be viewed primarily as a deployment consideration rather than a limitation of the execution-governance model itself.

Latency and Performance

A further concern is that mandatory validation before externally effective operations may introduce additional latency.

The architecture recognises that different processing activities have different assurance requirements. Controller-defined policies may therefore determine where protected validation is required and where conventional processing remains appropriate.

Furthermore, many GDPR-sensitive operations—such as cross-border transfers, disclosure of personal data, processing of sensitive information, or high-risk AI applications—already

involve policy evaluation, authentication, authorisation, encryption, logging, or auditing. Protected execution-governance introduces an additional validation stage for such operations, but this additional verification may be proportionate where stronger technical assurance is required.

The balance between performance and assurance is therefore an engineering and deployment decision rather than an inherent architectural constraint.

Relationship to Existing AI Scaling Models

Current large-scale AI platforms are primarily optimised for throughput, elasticity, and rapid data processing. By contrast, the proposed architecture prioritises controlled execution, policy verification, and protected authorisation before externally effective operations occur.

These objectives should not necessarily be viewed as conflicting. Rather, they represent different optimisation priorities. Existing systems optimise for computational efficiency, whereas execution-governance architectures optimise for verifiable policy enforcement and controlled consequence generation.

The appropriate balance between these objectives depends on the regulatory, commercial, and operational requirements of the deployment environment.

Potential Industry Perspectives

Large AI infrastructure providers may express concerns regarding additional engineering complexity, operational overhead, interoperability with existing systems, or effects on large-scale AI training efficiency.

Such considerations are expected whenever new infrastructure introduces stronger technical controls. Similar concerns accompanied the adoption of transport-layer encryption, hardware security modules, trusted execution environments, secure boot, and confidential computing technologies, all of which introduced additional implementation complexity but were progressively adopted where stronger security or regulatory assurance became necessary.

Accordingly, implementation complexity should be evaluated alongside the potential benefits of stronger technical enforcement, improved accountability, and reduced regulatory risk.

Incremental Deployment

The architecture is intended to support phased adoption rather than requiring immediate replacement of existing AI infrastructure.

Potential deployment pathways include:

- high-risk AI systems subject to enhanced regulatory obligations;

- processing involving special categories of personal data;
- cross-border data transfers;
- government and critical infrastructure systems;
- healthcare and financial applications;
- confidential enterprise AI environments;
- cloud services requiring stronger execution governance.

This enables organisations to adopt execution-governance selectively where additional technical assurance provides the greatest benefit.

Summary

The principal implementation challenges associated with execution-governance are engineering, operational, and economic considerations rather than fundamental technical impossibility. Legacy integration, latency, and deployment complexity are important practical factors that should be evaluated on a case-by-case basis. These considerations do not diminish the architectural objective of providing infrastructure-level technical enforcement for controller-defined processing conditions but instead inform appropriate deployment strategies and implementation choices.