

DEEP RESEARCH REPORT

Deep Research: Legal and Practical Imperatives for EU-Based Infrastructure

GENERATED	28 September 2026
REFERENCE	af45f439faed
PRODUCED BY	CIATA DeepResearch

This report was generated by AI (CIATA DeepResearch). It may contain errors or omissions. Verify critical facts independently.

Deep Research: Legal and Practical Imperatives for EU-Based Infrastructure

RESEARCH QUESTION:

What is "Deep Research", what are the legal and practical reasons to run it on EU infrastructure, and where do those reasons not apply?

Generated: 2026-09-28 17:11 UTC | Coverage: 0.90 | Iterations: 3 | Sources: 67 | Findings: 230 | Model: qwen3.8-27b

Executive Summary

Deep Research is an agentic AI capability that autonomously plans multi-step investigations, searches and reasons across heterogeneous web sources, and produces structured, citation-backed reports—designed to replicate the workflow of a human research team at machine speed [8]. The case for running such workloads on EU infrastructure is not a single legal mandate but the compounding effect of at least four overlapping instruments: the EU AI Act's risk-tiered obligations on model governance and transparency, the continued applicability of the General Data Protection Regulation (GDPR) alongside the AI Act's obligations, the European Health Data Space (EHDS) Regulation's channeling of secondary health-data use through in-EU secure processing environments, and the Digital Operational Resilience Act's resilience duties for financial-sector deployments [16] [29] [32] [34]. A structural sovereignty concern sharpens the argument: the U.S. CLOUD Act permits U.S. law-enforcement compulsion over data held by U.S.-headquartered companies regardless of server location, so a dataset in a Frankfurt region of a U.S. hyperscaler can remain reachable under U.S. jurisdiction, creating a conflict with GDPR Article 48 [16] [46]. The European Commission's AI Continent Action Plan frames this in strategic terms, emphasizing sovereignty and reduced dependency on external technologies [27].

Operationally, EU-hosted infrastructure eliminates 150–300 ms of intercontinental latency and \$0.02–\$0.09/GB egress charges that compound across the many retrieval and tool-call cycles in a single Deep Research session, while a growing set of compliance-native services—sovereign GPU clouds, data-residency-as-a-service platforms, and runtime telemetry that tracks every agent hop—reduces the engineering burden of meeting GDPR Article 32(1)(d)

continuous-evidence requirements [17] [18] [44]. Azure OpenAI, for instance, is stateless by design (prompts and completions are not retained or used for training) and encrypts data at rest by default, which simplifies the data-minimisation analysis for EU deployments [20].

The reasons for EU hosting do not apply uniformly. The AI Act's research exemptions in Articles 2(6) and 2(8)—covering systems built and used solely for scientific research, and research and development before a system is placed on the market—can relieve a qualifying project from the Act's broader obligations, though the GDPR continues to apply in full whenever personal data are processed [28] [30]. In the United States, the absence of a comprehensive federal AI statute means minimal AI-specific compliance overhead, though the emerging export restrictions on sensitive data categories introduce their own variables [10] [48]. China and Russia impose strict data-localisation regimes (as of August 2023) that are architecturally incompatible with EU-style Standard Contractual Clauses (SCCs) transfer flows, making them poor substitutes when a single coherent data-handling design is required [17] [42].

Deep Research: Scope, Architecture, and Distinguishing Features

Deep Research is an LLM-based AI process that uses tools to interact with the external environment in a multimodal and interactive manner with feedback, with the stated goal of assisting humans in discovering and solving problems at varying levels of automation and ultimately reaching or surpassing the capabilities of top human scientists [3]. In practice, it functions as an agentic capability that autonomously plans, searches, reasons, and reports on complex questions, synthesizing information from multiple web sources into a comprehensive report with verifiable citations [8]. The process is iterative: it semantically interprets a user's question, identifies and filters relevant content from various sources, and produces structured, well-founded summaries rather than a simple list of links [6]. Multiple rounds of searching, query refinement, and source comparison are driven by AI reasoning that determines subsequent information needs, making the workflow fundamentally different from a single-shot query [2]. The capability is designed to mimic the functions of a human researcher or team, performing deep dives into specific topics and providing detailed, real-world insights for tasks such as industry trend analysis, multidisciplinary knowledge synthesis, and argumentation [7].

The underlying architecture follows a structured sequence of distinct stages. One widely described formulation identifies four phases: planning objectives by breaking down the task, searching the web autonomously, reasoning across sources to identify patterns and resolve conflicts, and reporting findings with transparent source attribution [8]. A parallel description frames the same process as a structured sequence of planning, execution through multiple research turns, and a final synthesis phase to consolidate findings [4]. The methodology is designed for complex, multi-faceted inquiries—scientific interdisciplinary analysis, market research, policy analysis, and technical investigations—and employs adaptive planning and

progressive refinement to fill knowledge gaps rather than providing simple, single-response answers [4]. OpenAI's implementation, for example, is powered by a version of the o3 model optimized for web browsing and data analysis, utilizing reasoning capabilities to search, interpret, and analyze large volumes of text, images, and PDFs, and can be trained using reinforcement learning to execute multi-step trajectories [1]. That same implementation integrates chain-of-thought reasoning to generate intermediate "thinking tokens" for improved logical accuracy [8]. Google's Gemini deep research, by contrast, emphasizes structured planning and leverages Google's search infrastructure to provide broad source access, particularly for current events and trending topics [8].

Deep Research is distinguished from standard research tools and from related AI concepts in several specific ways. It is more automated and intelligent than standard research workflows; it incorporates search and review as key internal steps rather than being limited to them; and it differs from AI for Science (AI4S), which focuses on using AI as a tool to support specific scientific domains [3]. Compared with classic web searches, Deep Research provides direct, context-based answers rather than a list of links [6]. Unlike standard chatbots or search engines, it uses multi-step reasoning to synthesize information from multiple sources into a comprehensive, cited report [8]. It also differs from "Regular Ask" queries by employing adaptive planning and progressive refinement to fill knowledge gaps [4]. In the academic literature space, specialized tools such as Elicit, Undermind, SciSpace, and Ai2 Scholar QA use semantic search to find papers related to the meaning of a query rather than keyword matching, allowing synthesis across multiple studies [2].

The practical scope of Deep Research spans intensive knowledge work in finance, science, policy, and engineering, with the design intent of providing thorough, precise, and reliable research accompanied by fully documented outputs including clear citations and summaries of the reasoning process [1]. Sector applications include science and education for literature research, journalism for fact-checking, management consulting for market analysis, and law for reviewing legal texts and case law [6]. In scientific and pharmaceutical R&D, AI technologies of this kind are used to analyze biomedical data, identify drug candidates, predict efficacy, and optimize clinical trial designs [30]. Several platforms now offer Deep Research capabilities—Perplexity.ai for structured answers with source references, ChatGPT Pro for online research, Scite.ai for scientific citation analysis, and Gemini among others—with availability and usage limits varying by subscription plan [2][6]. Parallel provides an enterprise-grade Deep Research API designed for AI agents rather than consumer chat interfaces, returning structured JSON with verifiable citations and offering asynchronous processing via its Task API, with runtimes ranging from seconds to 45 minutes [8]. As of February 5, 2025, OpenAI's Deep Research was made available to Pro users in the United Kingdom, Switzerland, and the European Economic Area (EEA) [1].

Despite its comprehensiveness—outputs often exceeding 20 pages—Deep Research is not a replacement for human researchers. It can be prone to hallucinations, logical missteps, and lacks a default exploratory or creative mindset ^[7]. OpenAI's implementation, in particular, is noted for its high quality and range of use cases but is described as expensive ^[5]. These limitations define the technical baseline against which the regulatory, data-residency, and operational analyses in the sections that follow are measured: the system's autonomy, multi-source ingestion, and generation of structured outputs are precisely the characteristics that trigger obligations under EU law.

EU Regulatory Landscape for Deep Research: AI Act, GDPR, and Sector-Specific Instruments

The EU regulatory landscape governing Deep Research is not a single statute but a layered architecture in which the EU AI Act, the General Data Protection Regulation (GDPR), the European Health Data Space (EHDS) Regulation, and the Digital Operational Resilience Act (DORA) each impose distinct obligations that intersect and compound one another. The AI Act (Regulation (EU) 2024/1689) entered into force on 1 August 2024 and establishes a uniform, risk-based legal framework for the development, placement on the market, and use of AI systems within the EU ^[14]^[32]. The GDPR, fully enforceable since 25 May 2018, provides the foundational data-protection layer that the AI Act is expressly designed to complement rather than displace: the AI Act does not affect existing Union laws on personal data protection, including the GDPR itself ^[13]^[29]. The EHDS Regulation, which took effect on 26 March 2025, adds a specialised regime for health data that supplements and specifies GDPR requirements ^[34]^[35]. DORA, effective from January 2025, extends operational-resilience and data-transparency duties to financial institutions and their critical ICT providers ^[16]. Under this framework, many health- and pharmaceutical-related AI applications count as high-risk and require strict oversight of training-data provenance, model validation, and bias mitigation ^[10].

The EU AI Act: Risk Tiers, Obligations, and Enforcement Timeline

The AI Act categorises AI systems into four risk levels—unacceptable, high, transparency (limited), and minimal/no risk—and imposes obligations proportionate to each tier ^[9]^[12]. Prohibited practices, including social scoring, untargeted scraping of internet or CCTV imagery for facial recognition, and emotion recognition in workplaces, became effective in February 2025; the prohibition on non-consensual sexually explicit content generation takes effect in December 2026 ^[9]. High-risk systems—those deployed in critical infrastructure, education, employment, law enforcement, and analogous domains—face mandatory risk assessments, high-quality training datasets to minimise discrimination, activity logging for traceability, and elevated cybersecurity and robustness requirements, with strict obligations applying from 2 December 2027 ^[9]^[10]. Generative AI, including the model architectures

underlying Deep Research, is not classified as high-risk per se but must comply with transparency requirements: disclosing that content is AI-generated, implementing measures to prevent illegal content generation, and providing summaries of copyrighted data used for training ^[47].

The Act's implementation is phased. As of July 2024, the planned timeline was: bans on unacceptable-risk systems to take effect on 2 February 2025; transparency and copyright obligations for General Purpose AI (GPAI) models to begin in August 2025; and the majority of provisions, including obligations for most high-risk systems and mandatory conformity assessments, to become applicable by 2 August 2026 ^{[19][46][47]}. Article 4 requires organisations to ensure staff competence in handling AI systems (effective February 2025), while Articles 51 and following impose obligations on companies fine-tuning foundation models or building on GPAI systems (effective August 2025) ^[31]. Providers of GPAI models that pose systemic risks must assess and mitigate those risks and provide summaries of training content, including data sources ^[9]. The AI Office, part of the European Commission and supported by a Scientific Panel of independent experts, holds enforcement powers over GPAI models including the ability to request technical documentation and issue fines, with full enforcement commencing 2 August 2026 ^{[9][11]}.

The AI Act's territorial reach is broad. It applies to providers, deployers, importers, and distributors of AI systems within the Union market, including those located in third countries if the AI system's output is used in the Union ^{[12][29]}. It extends to any AI system that impacts individuals within the EU regardless of whether the provider or hosting infrastructure is located in the EU or a third country ^{[28][33]}. This extraterritorial scope creates a "Brussels effect" whereby non-EU researchers and providers must comply with EU regulations to maintain market access ^{[25][32]}. The Act explicitly excludes national security and defence, classified as a Member-State exclusive competence, though commentators warn this could create loopholes if AI applications are broadly characterised as national-security matters ^[25]. It also does not apply to public authorities in third countries using AI for international law-enforcement or judicial cooperation with the Union, provided adequate fundamental-rights safeguards exist ^[29].

Penalties are tiered. Violations of prohibition rules carry fines of up to €35 million or 7% of global annual turnover, whichever is higher; non-compliance regarding GPAI obligations can attract fines of up to €15 million or 3% of global turnover ^{[31][32][47]}. As of 2 August 2026, the transparency obligations are officially applicable, and non-compliance with them can result in fines reaching up to €15 million or 3% of a company's global annual turnover ^[24]. For AI researchers, the compliance burden is significant: if work falls under regulated categories, obligations include risk-management systems, data governance, and documentation of design choices, and for high-risk systems these obligations must be maintained throughout the

system's entire life cycle, including after publication [32]. To support implementation, each EU Member State is required under Article 57 to establish at least one AI regulatory sandbox by 2 August 2026, allowing controlled testing of AI systems within a regulatory framework [11] [27].

GDPR: Data-Protection Obligations Specific to AI Workflows

The GDPR applies to any organisation processing the personal data of individuals in the EU, regardless of the organisation's physical location, and has extra-territorial effect where organisations offer goods or services to, or monitor the behaviour of, data subjects within the Union [15] [26] [45]. Personal data is defined broadly to include any information relating to an identified or identifiable natural person—names, addresses, IP addresses, cookie data, genetic or biometric data [15]. The Regulation mandates core principles of lawfulness, fairness, and transparency; purpose limitation; data minimisation; accuracy; storage limitation; integrity and confidentiality; and accountability [13]. Processing is lawful only under specific Article 6 conditions: unambiguous consent, necessity for a contract, compliance with a legal obligation, protection of vital interests, performance of a task in the public interest, or pursuit of legitimate interests [13].

Four GDPR provisions are particularly critical for AI-agent compliance in a Deep Research context. Article 5(1)(c) requires data minimisation at inference; Article 30 requires continuously refreshed records of processing (ROPA) based on observed behaviour; Article 32(1)(d) demands ongoing effectiveness assessments of technical measures; and Chapter V (Articles 44–50) regulates international transfers triggered by every non-EU tool call or delegation [44]. The GDPR distinguishes between data controllers, who decide why and how personal data is processed, and data processors, such as cloud providers, who process data on behalf of a controller [13]. Organisations must implement appropriate technical and organisational measures, including two-factor authentication, end-to-end encryption, staff training, and access limitation [13]. The GDPR applies to the processing of personal data for statistical purposes as indicated by Recital 162 [26], and GDPR rules apply to personal data fed into AI models, with AI Act compliance intended to build upon these established data-protection practices [12].

When integrating Deep Research into corporate environments, companies must establish clear rules for sensitive content and maintain GDPR-compliant processes [6]. Non-compliance carries two tiers of administrative fines reaching up to €20 million or 4% of an organisation's global annual revenue, whichever is higher [13] [26] [42]. The practical severity of these penalties is illustrated by the \$1.3 billion fine issued to Meta for standard, routine data transfers that regulators deemed non-compliant, rather than a security breach [19].

The EHDS Regulation: Health-Data-Specific Obligations

The EHDS Regulation, which entered into force on 26 March 2025, is a digital infrastructure designed to standardise and exchange health data across EU borders to improve medical care and promote scientific research [35]. It establishes a legal framework for both primary use (immediate medical care) and secondary use (scientific research and health-policy purposes) [35]. Where electronic health data constitutes personal data, its access and processing remain subject to GDPR requirements; the EHDS introduces procedural and substantive safeguards to ensure GDPR compliance and the protection of data subjects' rights during secondary use [34].

The Regulation mandates that secondary use of health data by industry or research groups must occur through public or supervised infrastructure, shifting the paradigm from transferring raw datasets outside the EU to accessing data via secure processing environments (SPEs) within the EU [10]. The EHDS does not create a central European "data lake"; instead, it uses an application-based model where datasets are assembled per permit and accessed through SPEs, focusing on structured access to existing data rather than imposing new data-generation obligations [34]. Data users request access to electronic health data from data holders for specific secondary purposes—scientific research, public health, education—through Health Data Access Body (HDAB) intermediaries, and access must be conducted within SPEs meeting strict security requirements [34]. Secondary use for research and innovation is permitted on an opt-out principle: use is allowed unless an explicit objection is made, though the regulation explicitly prohibits using this data for advertising, marketing, or harmful products [35]. Individuals retain the right to opt out of secondary use, subject to limited Member-State exemptions; this right does not apply if the data holder cannot identify a natural person in a dataset, such as when data is pseudonymised and cannot be linked to identifiers in an opt-out list [34].

The EHDS mandates high technical and organisational security standards, including encryption and strict access controls, following Article 32 of the GDPR; the European Commission is tasked with defining specific security requirements by 26 March 2027 [35]. Implementation is staged: Member States must set up national contact points by 26 March 2027, primary-use provisions and EHR system compliance requirements apply from 26 March 2029, and genetic information will only be accessible for secondary use starting in 2031 [35]. Most secondary-use provisions are scheduled to apply from March 2029, with full implementation at the EU and Member-State level expected between 2027 and 2035 [34]. Obligations under the EHDS do not apply to health-data holders established in non-EU countries unless they have an established presence in the EU; for multinational companies, the entity controlling the means and purpose of processing is considered the controller under GDPR and is therefore bound by EHDS rules [34]. Violations can result in fines of up to €20 million or 4% of a company's global annual turnover in particularly serious cases [35].

DORA, the CDSM Directive, and Sector-Specific Instruments

DORA, effective from January 2025, applies to financial institutions and their critical ICT providers and mandates operational continuity, data transparency, and higher standards for cross-border data controls and third-party risk management ^[16]. For Deep Research workloads deployed in or serving the financial sector, DORA imposes an additional layer of resilience and transparency obligations on top of the AI Act and GDPR. The Digital Single Market (DSM) Directive provides copyright exceptions for Text and Data Mining (TDM) under Articles 3 and 4: Article 3 allows TDM for scientific research by research organisations and cultural heritage institutions, while Article 4 allows TDM by any party unless the rightsholder has expressly reserved rights via machine-readable means ^[12]. The EU Data Act, which began mandating the removal of data egress fees in 2025 with a goal of total elimination by 2027, is intended to reduce vendor lock-in and facilitate easier data movement between providers ^[46]. The Cyber Resilience Act sets vulnerability and incident reporting obligations under Article 14 to take effect on 11 September 2026 ^[24].

At the national level, the German Federal Data Protection Act (BDSG), in force since 25 May 2018, supplements the GDPR by filling regulatory gaps and adapting provisions to the German legal system, empowering German Data Protection Authorities to enforce both the GDPR and BDSG ^{[43][45]}. The BDSG mandates specific protections for sensitive personal data such as health or biometric information, requiring explicit consent and guaranteeing data-subject rights including access, rectification, erasure, restriction of processing, data portability, and objection ^[43]. Germany also maintains data-protection rules within sector-specific laws for the energy sector, financial trade, and healthcare services ^[45], and the Telecommunications-Digital-Services-Data Protection Act (TDDDG), in force since 14 May 2024, provides data-protection regulations for telecommunication and digital-services providers and transposes the ePrivacy Directive's cookie-consent requirement into German law ^[45].

How the Instruments Intersect

The AI Act and GDPR are designed to operate in concert: the AI Act's Article 12 logging requirements layer onto existing GDPR obligations for high-risk AI systems, mandating automatic event logging to ensure traceability of operations in convergence with GDPR's demand for continuous data-trajectory traceability ^[44]. The AI Act's ex ante risk-assessment logic mirrors the GDPR's preventive compliance model ^[25], and the AI Act is intended to complement existing frameworks including the GDPR, the Digital Services Act, and the Digital Markets Act ^[22]. The EHDS, in turn, specifies GDPR requirements for the health-data domain and channels secondary-use processing through SPEs that must satisfy both GDPR Article 32 security standards and AI Act data-governance obligations ^{[10][35]}. The EU AI Act reinforces data-governance and oversight requirements for AI systems, making it particularly relevant for organisations deploying AI on sensitive or regulated datasets ^[16].

△ A critical practical intersection arises from the AI Act's phased timeline and the GDPR's immediate enforceability. Since August 2026 the transparency obligations apply ^{[24][46]}, but the EHDS's most consequential secondary-use provisions do not apply until March 2029 ^[34]. A Deep Research project processing health data today must therefore comply with the full GDPR and AI Act regime while the EHDS's structured-access model remains in a preparatory phase, creating a window in which ad-hoc data transfers that would later be channelled through SPEs may still be permissible under GDPR ^{[10][35]}. The Digital Omnibus on AI, adopted in 2026, simplifies the rules and reduces compliance costs by postponing obligations for certain high-risk AI systems and removing registration requirements for systems performing narrow procedural or preparatory tasks, though critics argue these exemptions could still allow systems to impact rights in areas like asylum or taxation ^[48]. Multi-level governance complicates implementation further: national authorities must enforce EU-wide rules, which can lead to "gold-plating" where Member States extend EU powers, or fragmentation due to diverse approaches to designating market-surveillance and notifying authorities ^[48].

Data Residency, International Transfers, and Sovereignty: GDPR Transfer Rules, Schrems II, and the U.S. CLOUD Act

The General Data Protection Regulation (GDPR) does not impose a blanket data-residency mandate requiring all personal data to remain physically within the EU; instead, it strictly regulates the conditions under which such data may be transferred outside the European Economic Area (EEA) ^{[18][21][23]}. The practical consequence, however, is that the legal complexity of demonstrating adequate protection in a third country led many organizations (as of 2023) to keep EU data within EU borders by default ^[17]. For Deep Research workloads this distinction is critical because data residency—the physical location where data is stored—and data sovereignty—which legal jurisdiction governs that data—are not the same thing ^{[16][19]}. A dataset hosted in a Frankfurt data centre may satisfy residency yet remain subject to U.S. jurisdiction if the operating company is U.S.-headquartered, a gap the CLOUD Act exploits (analysed below) ^{[18][19]}. GDPR Articles 44–50 set out the transfer rules, and the regulation applies to the processing of EU residents' personal data regardless of where that processing takes place ^{[15][16]}.

Three principal mechanisms lawfully permit the transfer of personal data from the EEA to a third country: an adequacy decision issued by the European Commission, Standard Contractual Clauses (SCCs), and Binding Corporate Rules (BCRs) for intra-group transfers ^{[15][23][43]}. Additional derogations, such as explicit consent or the need to protect an individual's vital interests, allow transfers without a formal mechanism in narrow circumstances ^{[40][43]}. The table below summarises the key features of each mechanism.

Mechanism	Legal basis in GDPR	Authorisation required	Principal limitation
Adequacy Decision	Art. 45	European Commission assessment	Subject to periodic review and revocation (e.g., Privacy Shield invalidated in 2020) [15] [40]
Standard Contractual Clauses (SCCs)	Art. 46; Art. 28(7) for controller–processor modules [36]	None—pre-approved by the Commission [36]	Requires a case-by-case transfer risk assessment; does not guarantee adequate protection in the recipient country and may not address divergent legal frameworks such as government surveillance (as of June 2021) [37] [39]
Binding Corporate Rules (BCRs)	GDPR (BCR provisions)	DPA authorisation, typically 18–24 months [40]	Limited to intra-group transfers within a single corporate group [37]

The updated SCCs, published by the European Commission on 4 June 2021 and entering into force on 27 June 2021, were designed to align with the requirements of the Schrems II decision [\[39\]](#) [\[41\]](#). They adopt a modular approach in which parties select clauses corresponding to their roles—controller-to-controller, controller-to-processor, processor-to-processor, or processor-to-controller—and they integrate the processor obligations of GDPR Article 28 [\[38\]](#) [\[39\]](#). Organisations deploying SCCs must conduct transfer risk or impact assessments (TIAs) that evaluate whether the recipient country provides a level of protection comparable to the EEA, taking into account data categories, recipient types, and local laws [\[40\]](#). Under the SCCs, parties are prohibited from limiting their liability toward data subjects or to each other; any contractual cap on liability risks invalidating the clauses as a transfer tool [\[38\]](#). For transfers originating in the United Kingdom, the EU SCCs must be supplemented by the UK Addendum published by the UK Information Commissioner's Office, and transfers outside Switzerland require adaptations to comply with the Swiss Federal Act on Data Protection [\[38\]](#).

The Schrems II judgment, delivered by the Court of Justice of the European Union in 2020, invalidated the EU-U.S. Privacy Shield and established that transferring EU personal data to U.S.-based providers requires additional safeguards because U.S. surveillance laws can undermine GDPR protections [\[15\]](#) [\[16\]](#). The 2023 EU-U.S. Data Privacy Framework partially restored a transfer mechanism, but its long-term stability remained uncertain as of 2023 [\[17\]](#). One source reports that the European Data Protection Board (EDPB) has requested a review of the EU-U.S. Data Privacy Framework due to concerns regarding the independence of U.S.

authorities such as the FTC, a challenge that follows the U.S. Supreme Court decision in *Trump v. Slaughter* and impacts the adequacy decision underpinning transatlantic data transfers [24]. Adequacy decisions are inherently revocable: the EU's invalidation of the Privacy Shield in 2020 and the scheduled review of the United Kingdom's adequacy decision in June 2025 illustrate that adequacy decisions are not permanent [40].

The U.S. CLOUD Act (Clarifying Lawful Overseas Use of Data Act) allows U.S. federal law enforcement to compel U.S.-headquartered companies to provide access to data stored anywhere in the world, regardless of the physical location of the servers [16] [19] [46]. This creates what one analysis terms a "Sovereignty Gap": data stored in a European region of a U.S.-based hyperscaler remains subject to U.S. jurisdiction, potentially violating GDPR [19] [46]. The conflict is sharpened by GDPR Article 48, which prevents the transfer of data to third-country authorities unless a recognised international agreement exists; a U.S. court order issued under the CLOUD Act does not satisfy this requirement, leaving a provider legally compelled to disclose data under U.S. law while simultaneously obligated to protect it under GDPR [16] [46]. Even temporary storage of non-U.S. entities' data in U.S. data centres can trigger CLOUD Act jurisdiction [18].

These structural tensions are compounded by the non-deterministic nature of agentic AI workloads. Unlike traditional SaaS applications that make residency decisions at provisioning time, AI agents make residency decisions at inference time: data may exit a selected region through external tool calls, cross-region vector-database replicas, or sub-agent delegation [44]. One source reports that sovereign-cloud offerings such as AWS European Sovereign Cloud, Azure EU Sovereign Data Zones, and GCP Assured Workloads constrain data-at-rest location and operator nationality but are insufficient for AI agents, because they often fail to constrain the runtime trajectory and typically provide annual audits rather than the continuous evidence required by GDPR Article 32(1)(d) [44]. Azure OpenAI illustrates the deployment-level choice: "Global" deployments may process data in any geography where the model is deployed, whereas "DataZone" deployments restrict processing to the specified data zone, so a DataZone deployment in an EU Member Nation allows processing within any EU Member Nation [20]. AI-driven operations involving continuous data flows—sensor readings, model inferences, work orders—trigger regulatory implications across sovereign borders [19].

Beyond the EU-level framework, individual member states may impose additional residency requirements on specific sensitive sectors including health, finance, and defence [15] [18]. One source reports that German regulations, for example, require tax-related books to be kept within German territory and oblige telecommunications providers to store traffic data locally, although enforcement of the latter by the Federal Network Agency has been temporarily suspended due to legal proceedings [43]. Under GDPR Article 9, health, genetic, and biometric data are classified as special-category data requiring "essential equivalence" in protection levels for any third-party jurisdiction receiving them, and the EU is transitioning toward a federated data-infrastructure model that segments high-risk health data from non-EU data or

stores it locally within secure, compliant environments ^[10]. For organisations seeking to minimise transfer risk, practical options include using cloud providers with data centres exclusively located within the EU or implementing hybrid-cloud strategies that segregate EU data processing within EEA borders ^[15]. Where personal data is stripped of identifiers through PII masking or anonymisation at the source, it ceases to be personal data and the specific legal mandates governing personal-data residency no longer apply ^[17].

Operational Benefits of EU-Based Infrastructure for Deep Research

The practical case for locating Deep Research workloads on EU platforms rests on measurable performance gains, cost avoidance, and a maturing ecosystem of compliance-native services. Shipping data across continents introduces 150–300 ms of network latency and incurs cloud egress charges of \$0.02 to \$0.09 per GB, both of which compound across the many retrieval and tool-call cycles that characterise a Deep Research session ^[17]. Hosting data near the user base improves performance and sustains operational continuity, while also simplifying regulatory incident response during audits or breaches ^[18]. One source notes that local data centres within a given jurisdiction deliver improved performance through reduced latency and tighter control over data management, and that storing data within the EU can build customer trust by anchoring protection in European data laws ^[23].

On the security and compliance-by-design front, enterprise-grade Deep Research deployment demands data isolation, processing guarantees, and SOC 2 Type 2 certification to safeguard proprietary information ^[8]. Several vendors now package these guarantees into ready-made EU offerings. One source reports that Microsoft offers Sovereign Cloud options, a "Data Guardian" feature providing Europe-based access approval, and External Key Management, aligned to compliance frameworks including EU GDPR, EU Model Clauses, and ENISA IAF ^[20]. InCountry provides a cloud-based Data-Residency-as-a-Service that stores private data in Germany while permitting global access, layering data vaults, firewalls, encryption, and tokenization on top of SOC 2, PCI DSS, and HIPAA compliance ^[43]. For compute-intensive workloads, one source reports that Lyceum supplies sovereign GPU cloud infrastructure in which providers, hardware, and operational staff all remain within the EU/EFTA legal perimeter, functioning as a legal firewall against extraterritorial reach such as the U.S. CLOUD Act ^[46]. The same source adds that sovereign European GPU clouds are increasingly optimised for high-performance AI hardware including NVIDIA B200 and H100 systems, and that they avoid the jurisdictional conflicts faced by US-based hyperscalers operating in EU regions ^[46].

Privacy-enhancing technologies and runtime observability tools are extending these benefits into the agentic layer. One source reports that Duality Technologies offers a platform combining Fully Homomorphic Encryption and Federated Learning so that data analysis proceeds without ever exposing raw records, enabling cross-border collaboration without breaching residency or localisation rules ^[16]. As of May 2026, ARMO provides eBPF-based

telemetry and application-layer correlation that produce runtime-derived artifacts tracking the "data trajectory" across tool calls, retrievals, and delegations—including destination jurisdiction and processor identity—giving operators a verifiable audit trail for every hop a Deep Research agent makes [44].

These benefits come with trade-offs that temper the case for a purely EU-local stack. One source reports that complying with Data Residency laws can force investment in local cloud infrastructure or in-country data centres, and that using global providers such as AWS or Azure requires careful region-locking configuration that can itself increase costs and introduce latency [42]. Managing multi-region deployments to satisfy diverse national laws adds operational complexity, management overhead, and compliance cost for SaaS organisations [21]. In the public-sector context, a 2024 publication on AI in the public sector notes that AI uptake remains hindered by complex procurement processes, data-management difficulties, and a lack of regulatory clarity, with policy recommendations centred on promoting cross-border data sharing and clearer frameworks [14].

Research Exemptions and Scenarios Where EU Legal Requirements Do Not Apply

The EU AI Act carves out two distinct research exemptions, in Articles 2(8) and 2(6), that can relieve Deep Research projects from the Act's broader regulatory obligations. The first, the development-phase exemption, covers AI research, testing, or development activities conducted before a system is placed on the market or put into service, provided the work remains in a controlled laboratory setting and complies with applicable Union law [28] [29] [33].

The second, the scientific-use exemption, applies to AI systems or models that are both developed and put into service for the sole purpose of scientific research and development, a provision intended to protect scientific freedom and encourage experimentation and hypothesis testing [28] [30] [33]. In addition, the Act excludes AI systems used exclusively for military, defence, or national security purposes regardless of the performing entity, and it does not affect Member State competences in that domain [29]. The Act also provides exemptions for AI components distributed under open-source licenses and for activities carried out within regulatory sandboxes or real-life testing environments established by public authorities [12].

These exemptions are narrow and conditional, and several specific actions or circumstances cause them to lapse. Under the development-phase exemption, testing in real-world conditions—defined as temporary testing outside a laboratory using human participants—is explicitly excluded and triggers full compliance with the AI Act's strict rules, including prior approval, informed consent, and oversight [28] [33]. The scientific-use exemption is equally restrictive: a system that was originally developed for commercial or practical use, even if later deployed solely for research, does not qualify [33]. Similarly, if a hospital or other entity purchases a commercial AI tool and uses it for research, the tool does not benefit from the

exemption and remains subject to the Act ^[28]. The scientific-use exemption further expires the moment a system is "put into operation," a threshold reached when the system runs on real data or interacts with real users outside a controlled test environment ^[31]. A practical gotcha for research teams is that publishing model artifacts or code on platforms such as GitHub, Hugging Face, or Colab may be interpreted as "placing on the market" or "putting into service" an AI system, thereby voiding the research exception ^[32].

Even where a Deep Research project falls outside the AI Act's scope under Article 2(6), the General Data Protection Regulation (GDPR) continues to apply in full whenever personal data are processed, requiring compliance with standards for lawful, transparent, and purpose-limited processing and robust safeguards for individual privacy ^[30]. In the pharmaceutical sector, the European Federation of Pharmaceutical Industries and Associations (EFPIA) has specifically advocated that AI-based drug-development tools used solely for the research and development of medicines should be exempt from the AI Act's broader requirements, arguing that such obligations would stifle innovation ^[30]. The EU's Digital Omnibus on AI, a more recent legislative package, introduces additional regulatory relief: it provides exemptions for small mid-cap enterprises with up to 750 employees and permits the processing of special categories of personal data specifically for detecting and correcting AI bias, while simultaneously adding new prohibitions on AI systems that generate non-consensual intimate imagery or child sexual abuse material ^[24].

A persistent source of uncertainty across all these provisions is the absence of a harmonized definition of "scientific research" within EU law. This ambiguity is not confined to the AI Act; it extends to the GDPR and the EU Copyright Directive, making it difficult for Deep Research teams to determine with confidence whether their activities qualify for the relevant carve-outs ^[28]. Until the European Commission or the Court of Justice of the European Union provides a unified interpretation, projects that straddle the boundary between laboratory experimentation and real-world deployment—precisely the boundary that Deep Research workflows often cross—face a genuine risk that an exemption they relied upon will be found inapplicable.

Comparable Non-EU Regulatory Environments for Conducting Deep Research

The non-EU landscape for Deep Research workloads is best understood through three distinct regulatory paradigms: the EU's rights-driven model, the United States' market-driven model, and China's state-driven coordinated model ^[48]. Beyond these three, several jurisdictions—Russia, Australia, and the United Kingdom—impose data-residency or emerging AI-governance requirements that create additional compliance considerations for any Deep Research project seeking to operate outside the EU.

The United States operates in a market-driven regulatory environment characterized by the absence of a comprehensive federal AI framework and a reliance on various national bills and limited federal guidance [48]. As of December 2025, one source reports that the U.S. regulatory approach is decentralized and non-binding, relying on the NIST AI Risk Management Framework, with sectoral regulation through existing agencies for industries like healthcare, finance, and transportation [25]. Another source describes the U.S. approach to AI risk management as highly distributed across federal agencies, risk-based, and sectorally specific, relying heavily on non-regulatory infrastructure, existing federal agency authorities, and nonbinding guidance such as the Blueprint for an AI Bill of Rights [22]. On the data side, one source reports that the US follows a more open strategy for data use and global pooling of health data to drive AI innovation, contrasting with the EU's localization approach, though it has recently imposed restrictions on exporting certain data to 'foreign adversaries' like China and Russia [10]. In July 2025, the America's AI Action Plan was unveiled as a federal strategy designed to secure leadership in AI through three pillars: accelerating innovation via deregulation, building domestic infrastructure like data centers, and leading in international AI diplomacy and security [47]. One source characterizes the US approach to AI as a deregulatory, private-sector-led model aimed at achieving global technological dominance and national security, seeking to remove regulatory barriers and red tape to facilitate rapid infrastructure development [27]. For a Deep Research project, this means the United States offers minimal AI-specific compliance overhead but provides no equivalent of the EU's structured transfer mechanisms, and the emerging export restrictions on sensitive data categories could complicate cross-border data flows.

China presents the sharpest contrast to both the EU and the United States. It utilizes a state-driven coordinated model for AI governance, which serves as a third distinct regulatory paradigm [48]. The Personal Information Protection Law (PIPL) and Data Security Law impose strict localization requirements, mandating that personal information of Chinese citizens be stored within mainland China, with cross-border transfers involving more than 1 million individuals or 'important data' requiring security assessments by the Cyberspace Administration of China (CAC) [17]. China's Cybersecurity Law requires critical information infrastructure operators to store data within China and subjects any cross-border transfers to government scrutiny [18]. As of November 2021, one source reports that the PIPL is a comprehensive framework that requires local storage for Critical Information Infrastructure Operators (CIIOs) and mandates security assessments for cross-border data transfers [42]. China maintains strict data localization regimes through the PIPL, Cybersecurity Law, and Data Security Law, requiring data generated in China to be subject to local processing requirements and security assessments before any cross-border transfer [16]. The use of Standard Contractual Clauses (SCCs) in China is restricted to smaller transfers and requires registration with the cybersecurity authorities [40]. For a Deep Research project processing large volumes of personal data, China's regime is effectively incompatible with the EU-style SCC-based transfer architecture and would require a fundamentally different data-handling design.

Russia, Australia, and the United Kingdom each present distinct but narrower constraints. Russia's Federal Law No. 242-FZ mandates strict data localization, requiring that the personal data of Russian citizens be stored on servers physically located within Russia [17], and Russia mandates that all personal data of its citizens must be stored and processed on servers located within the country to comply with its sovereignty laws [18]. Australia enforces data residency in regulated industries via the Privacy Act and APRA regulations, which may require local hosting for certain data classes like tax or healthcare data [18]. In the United Kingdom, a private member's bill introduced in the House of Lords on 4 March 2025 proposes the establishment of a central AI Authority to oversee governance and ensure businesses adhere to safety and accountability principles [47].

Jurisdiction	Primary Data Residency Requirement	Cross-Border Transfer Mechanism
United States	No comprehensive localization; restrictions on exports to 'foreign adversaries'	Open data strategy; no SCCs equivalent
China	Strict (PIPL, DSL, Cybersecurity Law)	CAC security assessment; restricted SCCs with 10-day registration
Russia	Strict (242-FZ); servers physically in Russia	n/a
Australia	Sectoral (Privacy Act, APRA) for tax/healthcare data	n/a
United Kingdom	n/a	n/a

The table above summarizes the data-residency dimension, which is the most directly relevant constraint for Deep Research workloads that process personal data. The United States and the United Kingdom stand out as the only two jurisdictions in this comparison without a blanket localization mandate, though the U.S. export restrictions on sensitive data categories and the UK's proposed central AI Authority introduce their own compliance variables. China and Russia impose the most restrictive localization regimes, making them poor substitutes for EU-based infrastructure when the goal is to maintain a single, coherent data-handling architecture across multiple jurisdictions. Australia's sectoral approach is the least restrictive of the localization regimes but applies only to regulated industries, meaning a Deep Research project outside the healthcare or tax sectors may face no Australian-specific residency obligation at all.

Research Gaps

Several questions remain open because the available evidence does not resolve them:

- **Outcome of the EDPB review of the EU-US Data Privacy Framework.** One source reports that the European Data Protection Board has requested a review citing concerns about the independence of U.S. authorities such as the FTC, in the wake of the U.S. Supreme Court decision in *Trump v. Slaughter* [24]. Whether this review will lead to revocation of the adequacy decision—and what interim transfer mechanism would replace it—is undetermined.
- **Interpretation of "scientific research" under EU law.** The AI Act, the GDPR, and the EU Copyright Directive each reference the term without a harmonised definition [28]. Until the European Commission or the Court of Justice issues a unified interpretation, a Deep Research team cannot determine with certainty whether its activities fall within the Article 2(6) scientific-use exemption or trigger full compliance obligations.
- **Resolution of the CLOUD Act / GDPR Article 48 conflict.** No adjudicated mechanism or bilateral instrument currently resolves the situation in which a provider is simultaneously compelled by a U.S. court order and prohibited by GDPR from disclosing data to a third-country authority [46]. Whether future litigation or diplomatic negotiation will produce a workable rule is unknown.
- **Sufficiency of sovereign-cloud and runtime-observability solutions for agentic workloads.** One source reports that existing sovereign-cloud offerings constrain data-at-rest location and operator nationality but do not reliably constrain the runtime trajectory of an AI agent, and that they typically provide annual rather than continuous audit evidence [44]. Whether the emerging eBPF-based telemetry and application-layer correlation tools will mature into a generally accepted compliance standard is an open question.
- **Scope of the AI Act's national-security exclusion.** The Act excludes military, defence, and national-security applications as a Member-State competence [25], but commentators warn this could be exploited to characterise dual-use AI systems as national-security matters. How broadly national authorities will invoke this carve-out in practice is not yet established.
- **Status of the United Kingdom's proposed central AI Authority.** The bill introduced in the House of Lords in March 2025 remains at the private-member stage [47]; whether it will progress to enactment, and what governance structure it would ultimately create, is unresolved.
- **Resumption of German telecommunications data-storage enforcement.** One source reports that enforcement of the local traffic-data storage obligation by the Federal Network Agency has been temporarily suspended due to legal proceedings [43]. The timeline and conditions for resumption are not specified in the available evidence.

The evidence covers EHDS health-data obligations only shallowly. The report does not address quantitative latency benchmarks, detailed cost-benefit modelling or sector-specific implementation examples.

References

- [1] openai.com — Introducing deep research
- [2] ask.library.arizona.edu — What is Deep Research (in generative AI tools)?
- [3] arxiv.org — From Transformer to Agent, From AI to AI for Science
- [4] asyncfunc.mintlify.app — Deep research - Introduction to DeepWiki-Open
- [5] natesnewsletter.substack.com — 9 Uses of Deep Research: Unlocking Hidden Insign...
- [6] ki-company.ai — Deep Research: Understanding and Using AI Research
- [7] lumenova.ai — Deep Research Is Epic, But Only If You Know How to Use it
- [8] parallel.ai — What is deep research?
- [9] digital-strategy.ec.europa.eu — AI Act | Shaping Europe's digital future - Euro...
- [10] atlanticcouncil.org — Navigating the European Union's AI and health data ...
- [11] artificialintelligenceact.eu — EU Artificial Intelligence Act | Up-to-date deve...
- [12] sas-dhrh.github.io — EU Legal Framework | A toolkit for Generative AI, data ...
- [13] gdpr.eu — What is GDPR, the EU's new data protection law?
- [14] ec-europa-eu.libguides.com — EC Library Guides: Artificial Intelligence, Legal...
- [15] kiteworks.com — Understand and Adhere to GDPR Data Residency Requirements
- [16] dualitytech.com — Data Sovereignty vs. Data Residency: Full Comparison
- [17] expanso.io — Data Residency Requirements: A Complete Guide ... - Expanso
- [18] stonefly.com — Data Sovereignty Vs Data Residency - StoneFly, Inc
- [19] oxmaint.com — The Geopolitics of Data Residency: Navigating AI ...
- [20] learn.microsoft.com — Inquiry on Data Residency, Compliance, and Security for ...
- [21] payproglobal.com — What is Data Residency? Sovereignty & Compliance
- [22] brookings.edu — The EU and U.S. diverge on AI regulation: A transatlantic ...
- [23] billtrust.com — GDPR and Data Residency for International Firms in Europe
- [24] linkedin.com — AI Rules Kick In, the EU-US Data Privacy Framework Is ...
- [25] globalpi.org — AI Regulation Across the Atlantic: EU AI Act vs. U.S. ...
- [26] pro.bloomberglaw.com — Comparison Charts: U.S. State vs. EU Data Privacy Laws |...
- [27] dcnglobal.net — US vs EU AI Plans - A Comparative Analysis of the US and ...
- [28] nature.com — Challenges in applying the EU AI act research exemptions ...
- [29] artificialintelligenceact.eu — Article 2: Scope | EU Artificial Intelligence Act
- [30] clinicaltrialvanguard.com — AI in Research: Key Takeaways from the EU AI Act
- [31] linkedin.com — The Research Exemption in the EU AI Act Is Not What Most ...
- [32] arxiv.org — Position: EU AI Act's Research Exemptions Can Break ...
- [33] healthmanagement.org — EU AI Act Research Exemptions Under Scrutiny
- [34] goodlifesci.sidley.com — Navigating the European Health Data Space Regulation:...
- [35] activemind.legal — European Health Data Space and the EHDS Regulation
- [36] commission.europa.eu — New Standard Contractual Clauses - Questions and Answers...
- [37] kiteworks.com — Standard Contractual Clauses (SCCs): A Complete Guide
- [38] mayerbrown.com — European Commission's Q&A on the New Standard ...
- [39] tietosuoja.fi — Standard clauses adopted by the Commission
- [40] trustarc.com — Understanding Standard Contractual Clauses (SCCs)
- [41] developer.atlassian.com — Standard contractual clauses and cross-border transfe...
- [42] protecto.ai — What Is Data Residency? Requirements +Implementation ...

[43] incountry.com — Introducing Germany's data residency laws for global ...
[44] armosec.io — Privacy and Data Residency for AI Agents: What GDPR ...
[45] dlapiperdataprotection.com — Data protection laws in Germany
[46] lyceum.technology — EU Data Residency for AI Infrastructure: 2026 Guide
[47] metricstream.com — AI Regulation Trends 2026: Policies Across the US, UK &...
[48] theregreview.org — The Paradoxes of the European Union's AI Regulation

Methodology

This report was generated through 3 iteration(s) of autonomous deep research, using 67 source(s) and extracting 230 finding(s).

RESEARCH CONCLUSION

i Token budget exhausted — The research used the full token budget before reaching the coverage target. The topic may require deeper investigation — consider running a follow-up with a more focused question or different sub-topics.

ITERATION LOG

Iter	Coverage	Sources	Findings	IGR*	Duration
1	0.80	35	122	3.05	71.2s
2	0.90	20	67	2.23	45.5s
3	0.90	12	41	1.64	38.9s

*IGR (Information Gain Rate) = new findings per search result examined. Values near 0 indicate diminishing returns.

SOURCE DIVERSITY

- **Unique domains:** 44
- **Domain entropy:** 0.99 (0=mono-source, 1=perfectly diverse)
- **Dominant domain:** arxiv.org (4% of sources)

Data Processing Summary

AI Model: gemma4-chat-26b, gpt-oss-120b, qwen3.8-27b · AKI.io, Frankfurt, Germany

Storage: Hetzner, Germany

Search: Staan / SearXNG (EU)

Training: Your data is not used for model training

All data processed within the EU · GDPR-compliant · Deletable on request — privacy@ciata.io