VEIL.AI White Paper 2026:05

The end of the linear data pipeline

How privacy requirements are reshaping management of sensitive data for AI, research, innovation, and data-driven managament

veil.ai
ENABLING USE OF SENSITIVE DATA

For decades, enterprise data management followed a linear model. That model is no longer sufficient for sensitive data.

The Shift: Privacy requirements now dictate how data must be transformed before it can be used, re-used, shared, or commercialised.

The Challenge: Organisations must optimise privacy and utility simultaneously. Too little protection creates legal risk; too much destroys analytical value.

The Future: A fundamental shift from generic data preparation to use-case-driven, privacy-optimised data product creation. The era of raw data sharing as well as ready-made one-size-fits-all datasets is over.

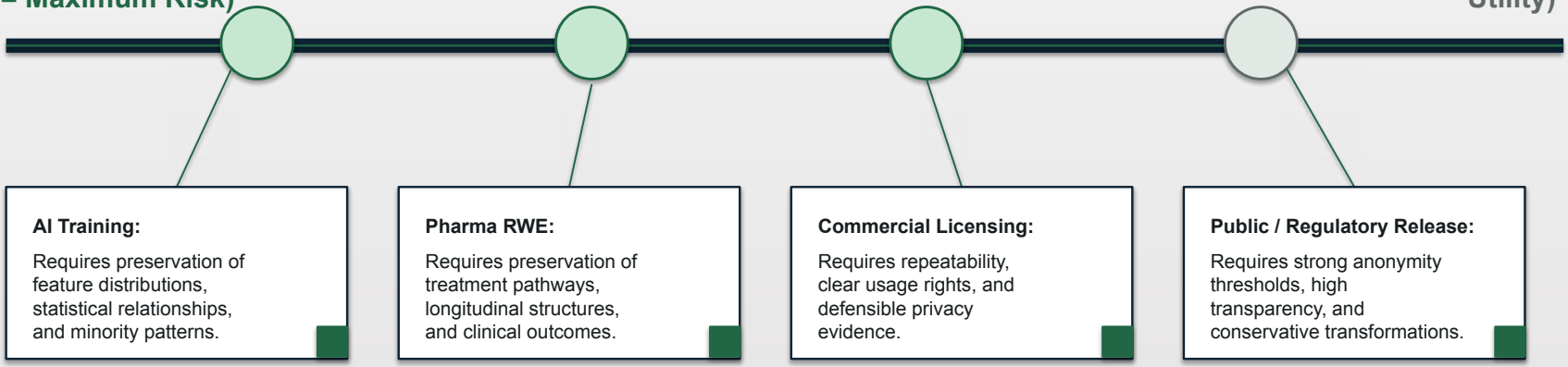
veil.ai

The Central Dilemma: Utility versus Privacy

Sensitive data is valuable because it is granular and context-rich.
However, the more useful the data is, the more sensitive it becomes.

High Data Utility (Raw/Pseudonymized Data = Maximum Risk)

100% Privacy (Destroyed Data = Zero Utility)



AI Training:
Requires preservation of feature distributions, statistical relationships, and minority patterns.

Pharma RWE:
Requires preservation of treatment pathways, longitudinal structures, and clinical outcomes.

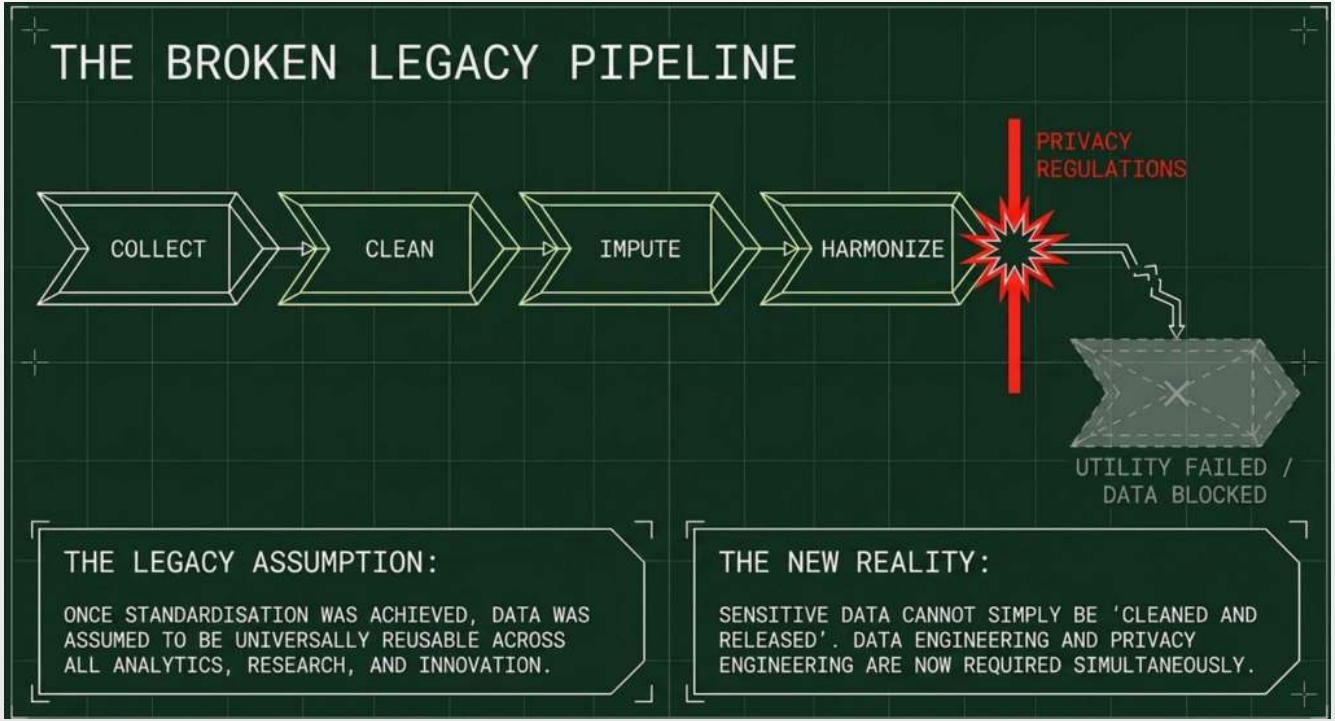
Commercial Licensing:
Requires repeatability, clear usage rights, and defensible privacy evidence.

Public / Regulatory Release:
Requires strong anonymity thresholds, high transparency, and conservative transformations.

There is no single universal version of "prepared data" anymore.
There are only fit-for-purpose, privacy-optimised data products.

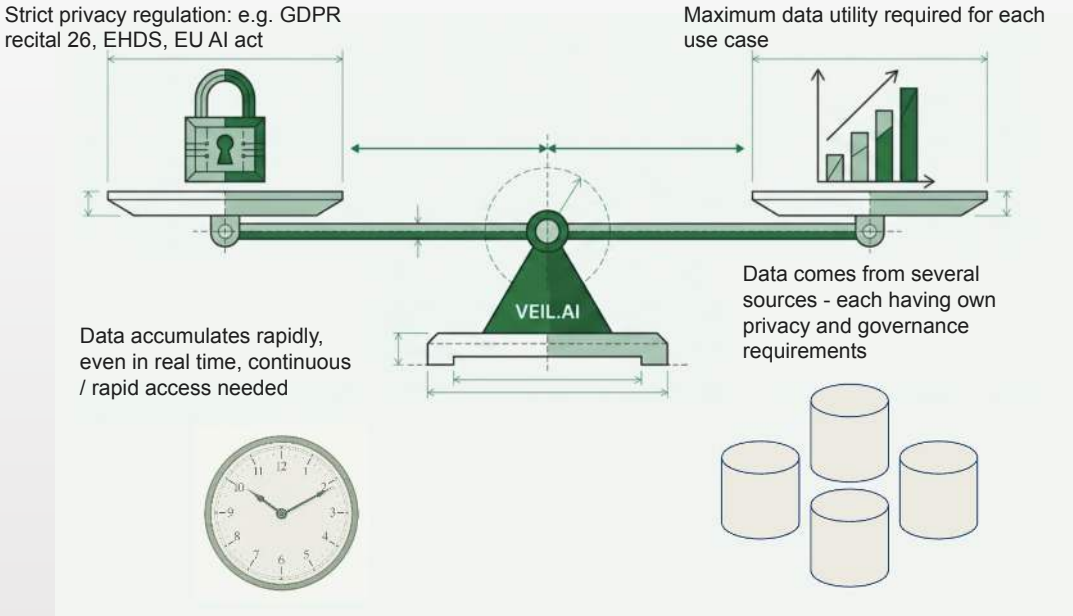
veil.ai

The Linear Data Pipeline Was Built For Consistency, Not Privacy



The Core Question Shift:
Old Paradigm: Is the data clean?
New Paradigm: Can this data be safely transformed into a privacy-preserving, high-utility asset for this specific use case? Can that be repeated for the next use case?

The new pipeline has to balance with privacy, utility, accumulating data, and multiple data sources



◎ Define use case
Clarify purpose, users, and decisions.

☰ Define privacy criteria
Identify legal, ethical, and risk constraints.

❙❙ Define utility targets
Specify analytic value, metrics, and thresholds.

⚙ Transform / optimize data
Apply techniques (e.g. Diff privacy, k, t or l diversity, synthetic data, etc)

⌕ Measure privacy risk and utility
Quantify risk and utility against chosen criteria.

↻ Iterate
Refine until the right balance is achieved.

☑ Release as privacy-optimized data product or otherwise
Fit for purpose.
Safe to share.

Continuous feedback to balance privacy and utility

veil.ai

Reversing the operational logic

When privacy matters, downstream end-user requirements must determine upstream data transformation.

The Old Logic:
Prepare Data First
→ Decide Uses Later

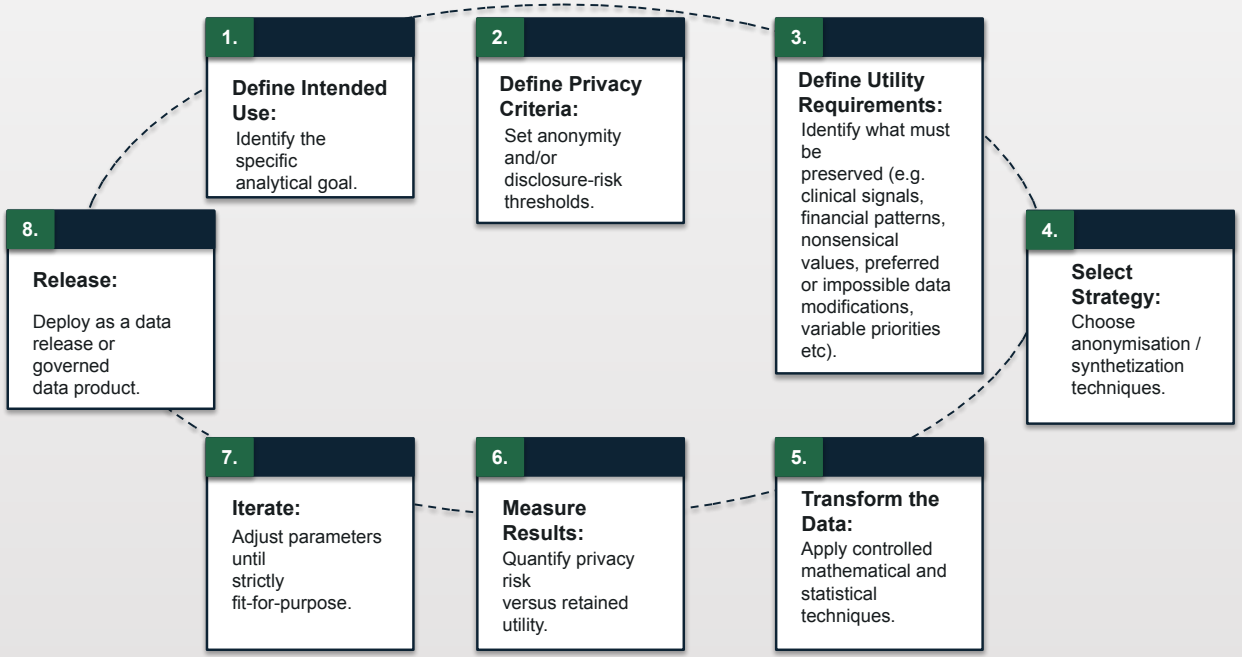
The New Logic:
Define Use Case First
→ Optimise Data Accordingly

Dimension	Legacy Model	New Model
The Trigger	Data collection	End-user requirement
Preprocessing Goal	Standardisation, cleansing, harmonization	Utility & Privacy Balance
Transformation Method	Generic (One-size-fits-all)	Tailored and utility & privacy engineered data, derived variables instead of original data, anonymization or synthetization
Final Output	A central Golden Record Database	Multiple, specific privacy-optimised data products (anonymized or synthetic). Detailed risk and utility assessments

The data flow is no longer linear. It is interlinked, iterative, and use-case-driven.

veil.ai

Managing sensitive data as a controlled transformation process



Key Insight

Privacy-preserving preparation is not a black box.

It requires deliberate, documented choices about what to preserve, what to alter, and what to Sacrifice.

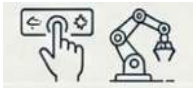
Well-done pipeline is transparent, provides documentation of what it does and provides comprehensive assessment of data utility and privacy

The VEIL.AI BONSAI Solution: Controlled, transparent, measurable, and scalable management of sensitive data

VEIL.AI provides the BONSAI platform and expertise layer to control use-case-driven optimisation, turning scarce data science expertise into institutional capability.

5. Institutional Codification

Capture project-level knowledge to create reusable workflows, agents, and skills.

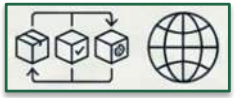


4. Progressive Automation

Apply expert oversight where required, and automate where processes become repeatable.

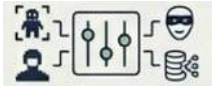
3. Privacy & Utility Preservation & Measurement

Preserve maximum useful information while mathematically proving privacy constraints are met.



2. Configurable Anonymity

Define specific privacy thresholds for internal and external AI, data or results sharing, or commercial data release.



1. Strategy Control

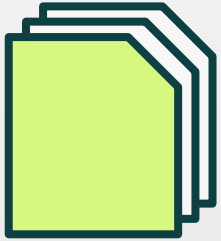
Define exactly how data elements should be anonymized, derived, engineered or , synthesized,

veil.ai

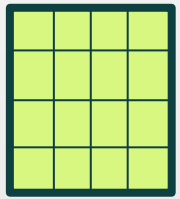
VEIL.AI BONSAI Anonymization & Synthetization Tools

BONSAI text

Free text
Accurate PII redaction
Black/white list
Bespoke language models



BONSAI structured



Tabular data
Row level
High utility
Comprehensive utility & privacy assessment

Customer hosted cloud or SaaS

Azure Managed App

Snowflake Native App

Professional services

- Complex anonymization/synthetization cases
- Detailed privacy risk analysis and reporting
- Customized enterprise solutions

From compliance burden to strategic asset portfolio

Advanced organisations do not merely share datasets; they engineer governed, documented, reusable privacy-wise secure anonymous data products.

The Anatomy of a Data Product:

☒ Defined Use Rights | ☒ Privacy Criteria | ☒ Utility Metrics |
☑ Transformation Logic | ☒ Versioning | ☒ Commercial Terms

Compliant Data Releases

Evidence-backed releases for legal, research, or regulatory partnerships (utilising anonymization, synthetization, data and privacy engineering , and utility & privacy assessments.

Data for AI Development

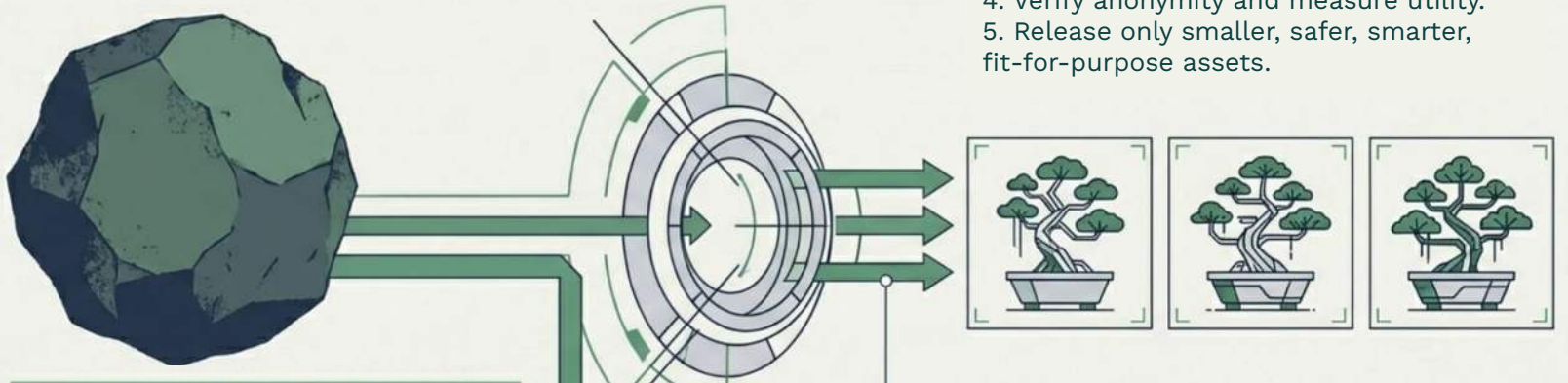
High-quality structured health, clinical text, or financial or signal data prepared safely for external LLMs or cloud platforms without sensitive data disclosure.

Monetisation & Commercialisation

Transforming restricted data liabilities into defensible, licensable commercial and intellectual assets.

The Strategic Imperative: Make Data A Critical Asset and Success Driver, Protect The Unique Data

The VEIL.AI BONSAI Data Strategy:
1. Protect the sensitive data strictly.
2. Define the precise intended use(s).
3. Control the anonymisation strategy.
4. Verify anonymity and measure utility.
5. Release only smaller, safer, smarter, fit-for-purpose assets.



Sensitive Data Liability: Sensitive data is your most valuable asset. Sharing it broadly, too early, or too cheaply with AI vendors or downstream users surrenders your control over privacy, competitive advantage, and future monetisation.

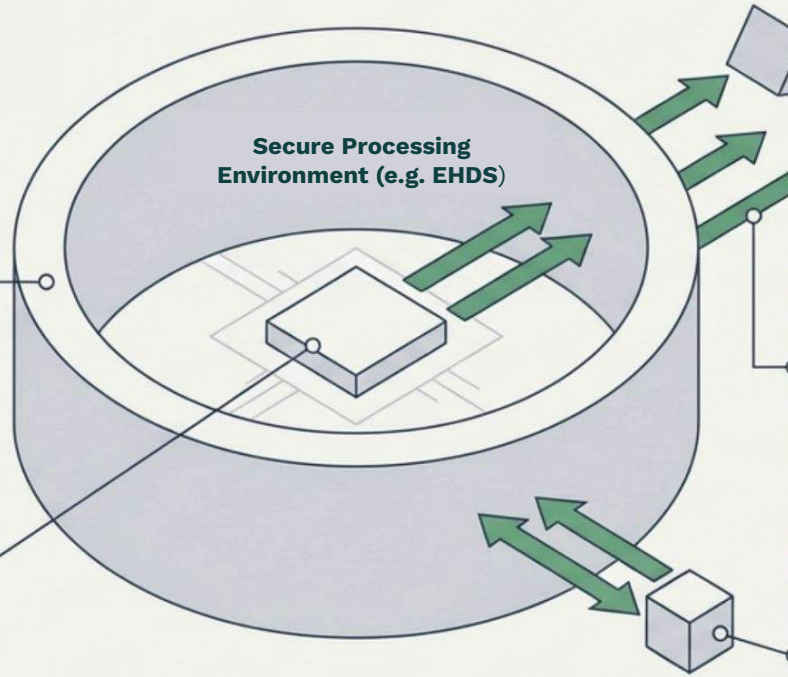
VEIL.AI BONSAI lens

The future is not sensitive data sharing. VEIL.AI enables the governed transformation of sensitive data into safe, useful, and even monetisable data products.

veil.ai

Special use case: securing privacy in federated architectures

Translating VEIL.AI capabilities into complex, secure, real-world environments.



A. Secure Environment Operations:

Operating entirely within the client's secure enclave.

B. AI Training Data Anonymisation:

Anonymising training data for demanding AI projects, ensuring models do not memorise or leak sensitive source information. Includes integrated risk and utility assessments.

C. Output Verification:

Rigorous verification and anonymisation of all results and data released from the secure environment.

D. Federated Analysis Assurance:

Guaranteeing anonymity and data safety across distributed, federated analytical networks.

veil.ai

Avoid overengineering: Start with one dataset. Prove value fast. Scale with confidence.

1 Recommended first step

1. Choose one high-value use case
AI training, research release, RWE analysis, test data, external collaboration, or data monetisation.

2. Run a BONSAI feasibility sprint
Assess privacy risk, utility requirements, transformation options, and expected data value.

3. Deliver a decision-ready output
A privacy-preserving sample dataset, utility/privacy report, and clear recommendation for scale-up.

2 What you get

A practical view of what can be preserved, transformed, or safely released

Evidence for legal, data, AI, and business stakeholders

A concrete **roadmap** from sensitive data liability to governed data asset

No need to commit to a full platform programme before the value is proven

3 Call to action

workshop → pilot → evidence → scale

Mailto: sales@veil.ai
juha.paakkola@veil.ai

veil.ai

ENABLING USE OF SENSITIVE DATA

Tuomo Pentikäinen

Chief Executive Officer

tuomo.pentikainen@veil.ai