CoreShield AI
SOVEREIGN AI PRIVACY GATEWAY

MainCore
TECHNOLOGIES

MAINCORE / SAIF 2026

CoreShield AI

The AI privacy gateway



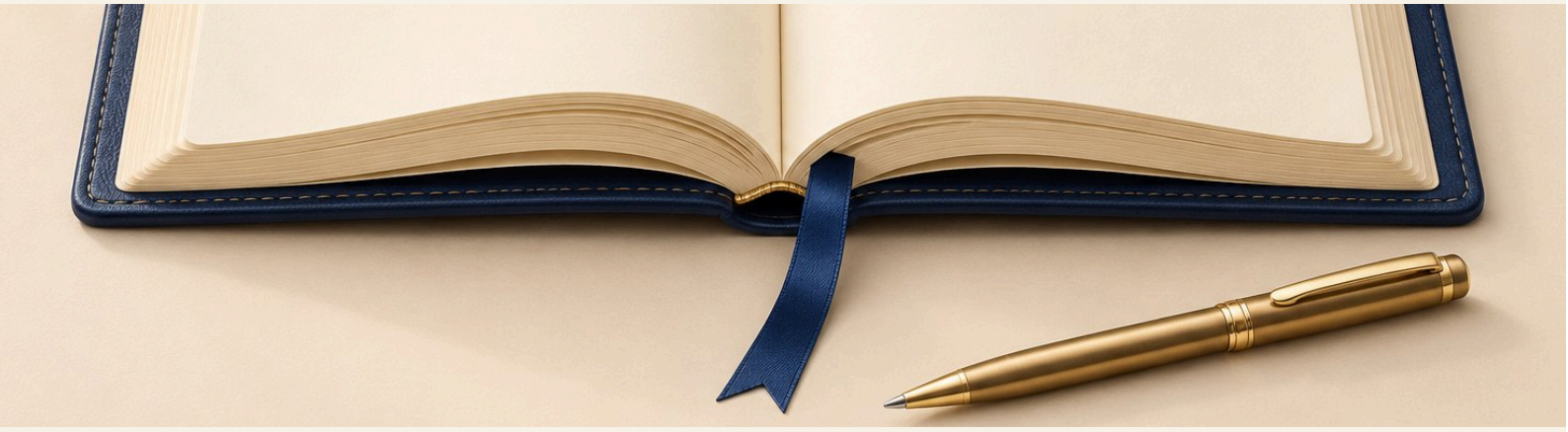
Between document and model

Visitor edition • SAIF, Riyadh

Software work registered with the Saudi Authority for Intellectual Property
No. 26-12-122599879 · 12/11/2026

FIELD GUIDE / 02

Your guide to this booklet



03–08 **Field guide**
From a contract on a desk to the trust boundary

09–16 **Context & use cases**
Saudi context, sectors and business functions

17–19 **Evidence & trust**
Measured figures, registration and safeguards

20 **Try it at the booth**
Demo steps and contact

Glossary



Detect
Finding the values a policy protects: names, identifiers, amounts and dates.



Token
A typed placeholder such as PERSON_TITLED_001 that replaces a value before text leaves.



Vault
An encrypted store that keeps each token's link to its original value, inside the organization.



Restore
Putting authorized original values back into the model's answer for the permitted user.



OCR
Reading text from scanned pages and images so their values can be detected too.



Precision & recall
Precision: share of flagged items that were truly sensitive. Recall: share of sensitive items that were found.

FIELD GUIDE / 03

A contract on a desk



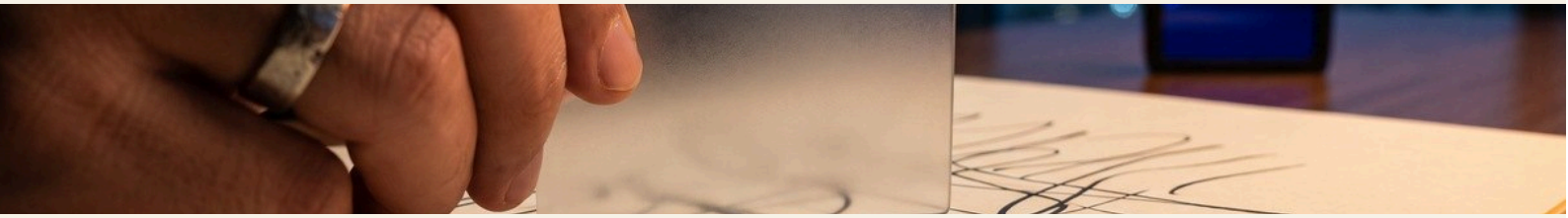
An employee needs a summary of a payment clause. The clause names a beneficiary, an amount and a due date. Pasting the whole clause into an external model sends those values along with it.

Why not simply delete the names?

If both names disappear, it can be hard to tell who pays whom. A consistent token gives each party a distinct reference within the request.

FIELD GUIDE / 04

The round trip



The document passes through four stages. The mapping between token and original value remains inside the trusted environment.



01

Detect

Identify entities covered by policy: a name, identifier, amount or date.



02

Mask

Replace each value with a typed token. Store its mapping in the organization’s encrypted vault.



03

Send

Apply policy, then send tokenized text to the model. The mapping table is never part of the request.



04

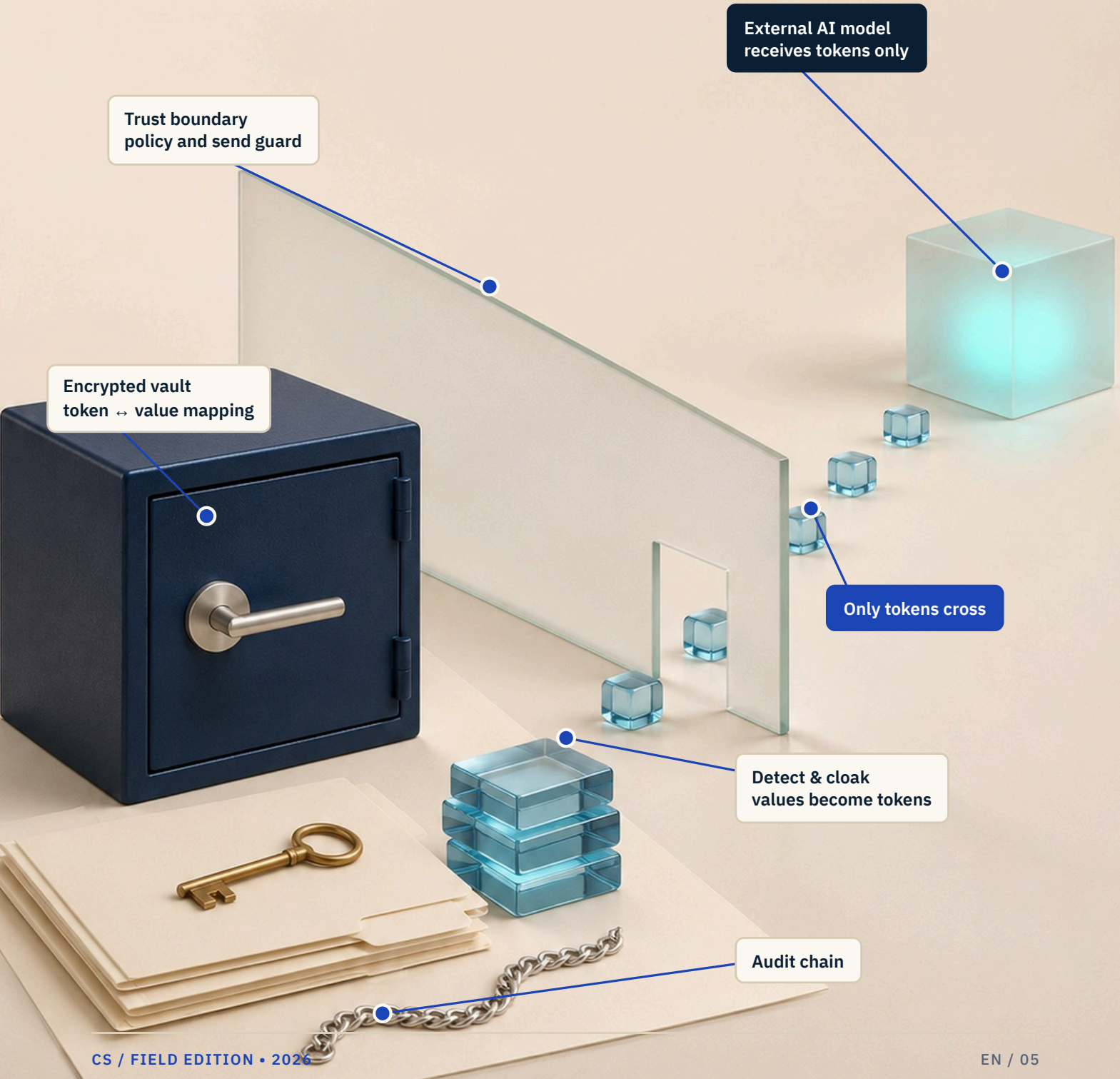
Restore

The gateway receives the response and restores authorized values for the user inside the trusted environment.

FIELD GUIDE / 05

Where does your data live?

Original values, the mapping table and the audit trail stay inside the organization. Only tokens cross the boundary.



FIELD GUIDE / 06

Three values in one clause

In the poster example, a person, amount and date can be traced from the original through the tokenized request to the restored response. All values are synthetic.



PERSON

Faisal Al-Dosari

PERSON_TITLED_001



AMOUNT

USD 4,500,000

AMOUNT_CURRENCY_001



DATE

15 Mar 2027

DATE_GREGORIAN_001

After processing, the authorized user sees the original values again. The model provider does not need the mapping table to produce the summary.

FIELD GUIDE / 07

Through the model’s eyes

Worked illustration using the poster’s values, not a transcript of a live session. Task: summarize a payment obligation.



Original → Tokens → Restored

Request inside the organization

Summarize this obligation: pay Faisal Al-Dosari USD 4,500,000 on 15 March 2027.

Values represented at the model

```
Pay PERSON_TITLED_001 the sum AMOUNT_CURRENCY_001 on DATE_GREGORIAN_001.
```

Summary after restoration

Faisal Al-Dosari is due USD 4,500,000 on 15 March 2027.

FIELD GUIDE / 08

The trust boundary





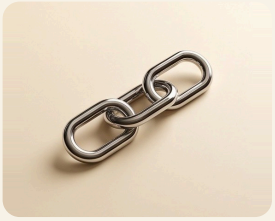
The vault

The token mapping is separate from model input, encrypted and scoped to each organization.



Authorization

Restoration is governed by policy. Possessing a token alone does not authorize access to its original value.



Audit trail

Processing decisions leave a trail that can be reviewed after model use.

FIELD GUIDE / 09

Arabic and Saudi context

The project’s rules address entities common in Saudi organizational records. The entity type carries meaning into the token; the original value stays in the mapping table.



Identity

National identity and residency identifiers.



Contact & banking

Local mobile numbers and Saudi IBANs.



Parties & transactions

People and organizations, monetary amounts and dates.

A rule does not guarantee detection of every format. Testing representative texts reveals what the rules miss.

FIELD GUIDE / 10

Where masking helps



Legal contracts

Summarize two parties' obligations while keeping the parties distinct without sending their real names.



Government correspondence

Classify the subject of a letter after masking the sender's name or national ID under organizational policy.



Financial records

Analyze a transaction description after masking its account or amount where policy permits.



Enterprise reporting

Draft a management summary while replacing sensitive names and keeping repeated references consistent.

Illustrative use cases, not customer deployments.

FIELD GUIDE / 11

People & essential services

Potential applications, subject to organizational policy



Healthcare

Summarize an administrative note or referral after masking patient identifiers, with human review of clinical content.



Insurance

Organize claim narratives while masking policyholder and policy identifiers; specialists retain decision authority.



Education & research

Summarize student administration or research interviews after treating names and identifiers for the stated purpose.



Nonprofits

Classify beneficiary requests and summarize programs while protecting contact and donation details.

FIELD GUIDE / 12

Economy & infrastructure

Potential applications, subject to organizational policy



Energy & utilities

Summarize service reports after masking customer and site details; review sensitive operational content separately.



Industry & mining

Organize quality and supplier reports while treating names, contract references and sensitive commercial details.



Real estate & construction

Summarize leases and project contracts while tokenizing parties and financial details under policy.



Transport & logistics

Classify shipment queries while masking sender, recipient, contact details and shipment references.

FIELD GUIDE / 13

Services & digital markets

Potential applications, subject to organizational policy



Retail & e-commerce

Summarize order complaints and customer feedback after treating names, payment and contact details.



Telecom & technology

Organize support tickets while masking subscriber details; technical secrets require tailored rules.



Tourism & hospitality

Summarize guest requests and bookings after masking identity, contact and reservation identifiers.



Media & professional services

Draft summaries of interviews or client files while protecting source identities and confidential details.

FIELD GUIDE / 14

Institutions & public trust

Potential applications, subject to organizational policy



Government services

Organize correspondence and service requests with masked identities; the agency controls what may be sent.



Banking & fintech

Summarize customer-service files after tokenizing accounts and identities; test the effect on meaning.



Legal & advisory

Review contract text while keeping parties distinguishable and protecting names and case references.



Security-sensitive services

Organize only permitted correspondence; classified material stays outside external models under policy.

FIELD GUIDE / 15

Across business functions

Sensitive data travels across departments. Detection, tokenization and restoration can support multiple workflows, with a policy tailored to each purpose.



Human resources

Summarize CVs and requests after masking personal details.



Procurement & supply chain

Compare proposal text while protecting suppliers and sensitive terms.



Finance & administration

Prepare summaries with control over accounts and amounts sent.



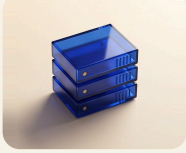
Customer service

Classify complaints and draft replies without exposing original identities.



Legal & assurance

Summarize documents and review processing decisions and audit trails.



Knowledge & technology

Process authorized text before it is used with models.

FIELD GUIDE / 16

From potential to deployment



01 Define data and purpose

Select permitted text and the fields requiring protection.

02 Set policy and permissions

Define what is tokenized, what cannot be sent and who may restore values.

03 Test representative cases

Measure missed entities, contextual leakage, restoration and response quality.

04 Expand in controlled stages

Review audit records and integration limits before adding departments and data.

These scenarios do not establish support for every format or integration, or prove compliance. Scope must be agreed and tested with each organization.

EVIDENCE / 17

What do the figures show?

32 deployed microservices, all ready	**257,000+** audit events, chain verified intact
119,786 sensitive values cloaked	**2,294** synthetic documents processed

Platform counts measured live on 16 Nov 2026. They describe the operating prototype; they do not measure detection accuracy or leakage prevention, which appear on the next page.

What should be tested next?

Independent evaluation on new data, Arabic compound names, Latin names written as ordinary words, low-quality scanned pages, adversarial inputs and latency under load. Reversible tokenization does not necessarily make text anonymous.

Source: live platform counters on coreshield.dev, 16 Nov 2026; CoreShield AI scientific poster, SAIF 2026.

EVIDENCE / 18

Tested at full document scale

Internal tests on synthetic data, 12–16 Nov 2026. Every figure here is a measurement, not an estimate.

200

PAGES IN ONE DOCUMENT

60,700 words and 6,000 sensitive values. After cloaking: zero raw values left, and 6,000 of 6,000 values restored identically in 5.4 s.

0

original values reached the model

A real external model (gpt-4o-mini) received the 200-page PDF: 447,977 characters, with no original value among them.

16

supported file types

PDF including scanned pages (OCR), Word, Excel, PowerPoint, images, HTML, CSV, JSON and text, plus audio and video, transcribed inside the platform; no raw audio leaves.

99.0%

of structured identifiers detected

National ID, Iqama, CR, IBAN and email, on a held-out test set (540 documents, 32,754 values). Phones 96.4%, person names 95.3%, organisations 85.3%; precision 97.8%.

0 / 140

raw Latin names passed the guard

In a replay of 140 test documents (73 before the fix), with none of 720 clean documents wrongly refused.

Stated limits: scanned pages depend on OCR quality, some compound and uncommon names are still missed, and employee numbers are at 61.9% and being improved. Measurements are internal and not yet reproduced independently.

TRUST / 19

Registered and built for trust



Software work registered with the Saudi Authority for Intellectual Property (SAIP)

REGISTRATION NO.
26-12-122599879

REGISTERED ON
12 Nov 2026

Tamper-evident audit log

SHA-256 hash-chained, append-only events; the database itself refuses any edit or deletion, and the chains are checked continuously.

Each organization isolated

Row-level security inside the database keeps every organization's data apart.

Encrypted vault

Every original value is sealed with AES-256-GCM envelope encryption.

Fail-safe by design

If detection, a scanned page or the send check cannot finish, the document stops and the user is told why; nothing is sent incomplete.

Arabic-first and bilingual

Arabic and English interface, right-to-left layout, and Arabic text read in its correct order from PDFs and scans.

Saudi regulatory mapping documented

A personal-data protection impact assessment and NCA control mappings (ECC, CCC): documented mappings, not certifications.

Independent research evaluation

A Saudi researcher specializing in Arabic-language AI is currently conducting an independent research evaluation of the platform.



Bring a question from your work

Choose a synthetic example resembling your organization’s documents. Follow what changes before the request is sent and what returns after restoration.

01 Pick a synthetic document close to your work.

02 Watch Detect and Cloak mark each sensitive value.

03 See exactly what the external model receives.

04 Restore the answer and compare it with the original.

SAIF project team / MainCore Technologies

Khalid Fahad Al-Turki
Founder & CEO

Musab Fahad Al-Turki
Co-Founder & Cybersecurity Lead

coreshield.dev

+966 56 936 6661

info@maincore.sa

maincore.sa

Scan to visit the platform