

DenseSense

Find where your controlled work actually sits on disk -- before you tell an assessor where it is. DenseSense answers the scoping question from the disk instead of from memory. Drive it from the ForteFide dashboard or the API. This guide is both the fast path (Section 2, first scan in about ten minutes) and the depth behind it. An Early-Adopter capability inside ForteFide, in active development.

- CUI marking discovery | Early Adopter (in active development)

- Four confidence tiers, never pooled into one number

- Paths + counts + tier only -- never the matched content

- Marking-focused -- not a PII/PCI detector

Early Adopter status -- read this first

STATUS

DenseSense is an **Early-Adopter** capability, in active development. The engine, the operator UI, and the API are real and tested; it is offered to early adopters while its taxonomy and operator surface keep tightening. [Contact us for early access.](#)

WHAT IS REAL

The detection engine (four tiers), the stdlib text extractor, the teach-your-own-forms path, the V2 dashboard screens, and the `/api/densesense/*` routes are built and tested.

WHAT IS NOT

It is **CUI-marking-focused** -- not a PII, PHI, or cardholder-data detector -- and it does not read text inside scanned images (no OCR). Those are stated gaps, not silent misses.

HOW TO REACH IT

In the dashboard it is the Step-2 setup and the Step-4 intel modal (an overlay, gated on the DenseSense entitlement); scripted, it is the `/api/densesense/*` API. On a build where the capability is not enabled, both report it as unavailable rather than half-running -- [contact us for early access.](#)

CONTENTS

- 1 The question DenseSense answers
- 2 Quick start -- your first scan in about ten minutes
- 3 Early Adopter scope -- what it is and is not
- 4 The four tiers -- graded, never pooled
- 5 What it reads -- and the one gap it names
- 6 Teach it your controlled forms (the DECLARED tier)
- 7 What the report says -- container, not content
- 8 The API surface
- 9 Where it fits in compliance
- 10 Early Adopter limitations & what is not built

1 | The question DenseSense answers

Every CMMC scoping decision starts with a claim about *where the controlled work is*. That claim is a guess until someone checks the disk. **DenseSense checks the disk.** Point it at your file roots and it reports where evidence of CUI actually sits -- which shares, which folders, and by what kind of signal -- so the boundary you draw for an assessor comes from findings, not from memory.

It is deliberately narrow about what it claims. DenseSense finds the **marking** -- the banner, the portion mark, the distribution statement, the export-control language -- because a marking is the one high-confidence signal that exists: the owner already told you the document is controlled. Everything softer than that is graded lower and labeled as such.

Tip -- CUI is not a data *format*. There is no regex for "a controlled part drawing." An engineering drawing, a tolerance spreadsheet, and a statement-of-work PDF can all be CUI while sharing no lexical feature at all. DenseSense detects the *designation's evidence*, and is honest about the confidence of each kind.

2 | Quick start -- your first scan in about ten minutes

DenseSense runs inside ForteFide. In the dashboard it is the Step-2 setup (point it at your shares and the customers who send you government work), a run, and the Step-4 intel modal that shows the result. The same loop is scriptable over the API, which is what the rest of this section shows. Every `/api/*` call carries the dashboard token as the `X-Auth-Token` header, so grab it once:

```
TOKEN=$(sudo cat /var/lib/fortefide/dashboard.token)
```

1. Confirm it is enabled. A build with DenseSense enabled reports `densesense` on; a build without it reports it off, and both the dashboard tile and the scan route report it as unavailable rather than a half-run.

```
curl -s http://127.0.0.1:5000/api/capabilities \
  -H "X-Auth-Token: $TOKEN" | grep densesense
```

2. Start a scan over the roots that actually hold your work:

```
curl -s -X POST http://127.0.0.1:5000/api/densesense/scan \
  -H "X-Auth-Token: $TOKEN" -H "Content-Type: application/json" \
  -d '{"roots": ["/srv/engineering", "/home/shared"]}'
# -> 202 {"job_id": 1}
```

It returns `202` and a `job_id`. An empty `roots` is refused with `400 roots_required` -- it never scans a scope you did not state.

3. Read the result -- poll the job until it completes:

```
curl -s http://127.0.0.1:5000/api/densesense/scan/1 \
  -H "X-Auth-Token: $TOKEN"
```

The result carries each tier's count *separately* (Section 4), the per-file findings -- path, marking type, tier, and count, **never content** -- and the heatmap and concentration that show where the controlled work piles up (Section 7).

Tip -- That is the entire fast path. Everything below is the depth: what each tier means, what DenseSense can and cannot read, how to teach it your unmarked forms, and how to read the report without putting CUI into it.

3 | Early Adopter scope -- what it is and is not

DenseSense is an **Early-Adopter capability, in active development**. The engine, the V2 operator screens, and the API are real and tested; the honest limits are about *what it detects*, not whether it exists. Read this table as the boundary of the claim.

IS	IS NOT (TODAY)
A CUI- marking discovery engine with four graded confidence tiers	A PII, PHI, or cardholder-data (PCI) detector -- those regimes are out of scope
Driven from the ForteFide V2 dashboard (a setup step and an intel modal) or the API	An OCR engine -- a marking printed inside a scanned image cannot be seen without OCR, and that is a stated gap, not a silent miss
On-host and self-contained: it walks roots you supply and keeps nothing it reads	A cloud service -- there is no upload, no phone-home, no telemetry

In the dashboard, DenseSense is a Step-2 setup: point it at the shares that hold your work and name the customers who send you government work. It is an overlay on the six-step flow, gated on the license -- not a seventh step.

DENSEDEFENSEFORTEFIDE

1 - FIND

2 - SET UP

3 - CHECK

4 - RESULTS

5 - FIX

6 - REPORT

LICENSE

SIGN IN SO WE CAN SET UP YOUR COMPUTERS

To set up the secure login on each computer, ForteFide needs an **administrator account** for them — one that can install software and set up a login. Enter it below. It's used only to set things up, and everything it creates is removed when you're done.

TYPE OF SECURE LOGIN TO SET UP

Certificate (recommended)

LINUX (SSH) ADMIN USERNAME

root or admin user

LINUX (SSH) ADMIN PASSWORD

Password

WINDOWS ADMIN USERNAME

Administrator

WINDOWS ADMIN PASSWORD

Password

What permissions does it need?

Enter an administrator account so we can set up your computers.

Do you know where you keep your government work?

Yes — I'll show you the folders

Not really — go look for it

Which of your customers send you government work?

Peach Cobbler

+ add another

Most shops name three or four. You can add more later.

STEP 2 OF 6 - SET UP

We found 3 computers.

Before we check them, ForteFide sets up a temporary, secure login on each one. It's removed automatically when you're finished — nothing is left behind.

214 (192.168.30.214)

a Linux computer

20 (192.168.30.20)

a Windows computer

DC01 (192.168.30.100)

your main Windows server

Ask a question

private · on this pc

Step 2 of 6

DenseSense setup in the ForteFide dashboard: point it at scope and at the customers who send you controlled work. Gated on the DenseSense license; an overlay, not a new step.

4 | The four tiers -- graded, never pooled

A discovery report that collapses everything into one number is a report the customer mutes. DenseSense grades every finding and **reports the tiers separately -- there is deliberately no grand total**. Strongest first:

TIER	WHAT IT MEANS	BASIS
CONFIRMED	A CUI marking is present. The owner already told you.	Banner / portion markings: CUI , CUI//SP-CTI , legacy FOUO
DECLARED	Matches a document <i>you</i> designated as controlled -- the only source of truth for CUI that carries no marking at all.	Your taught forms (Section 6), matched by cui_rules , provenance kept to the report
PROBABLE	A strong inference worth treating as a finding.	Distribution statements (B-F), export-control language (ITAR / EAR / ECCN)
POSSIBLE	A lead worth a human look. Not a finding.	A weak pattern with no corroboration -- surfaced for triage, never counted as CUI

Note -- The tiers are never merged. A CONFIRMED marking and a POSSIBLE lead mean very different things to an assessor, and pooling them into "1,482 hits" destroys the exact distinction that makes the report usable. summary() returns each tier's count on its own.

In the operator dashboard these same grades appear as plain-language urgency bands -- *deal with first, review soon, note for later* -- so a shop owner reads a priority, not a taxonomy; the underlying tier is preserved for the report and the evidence.

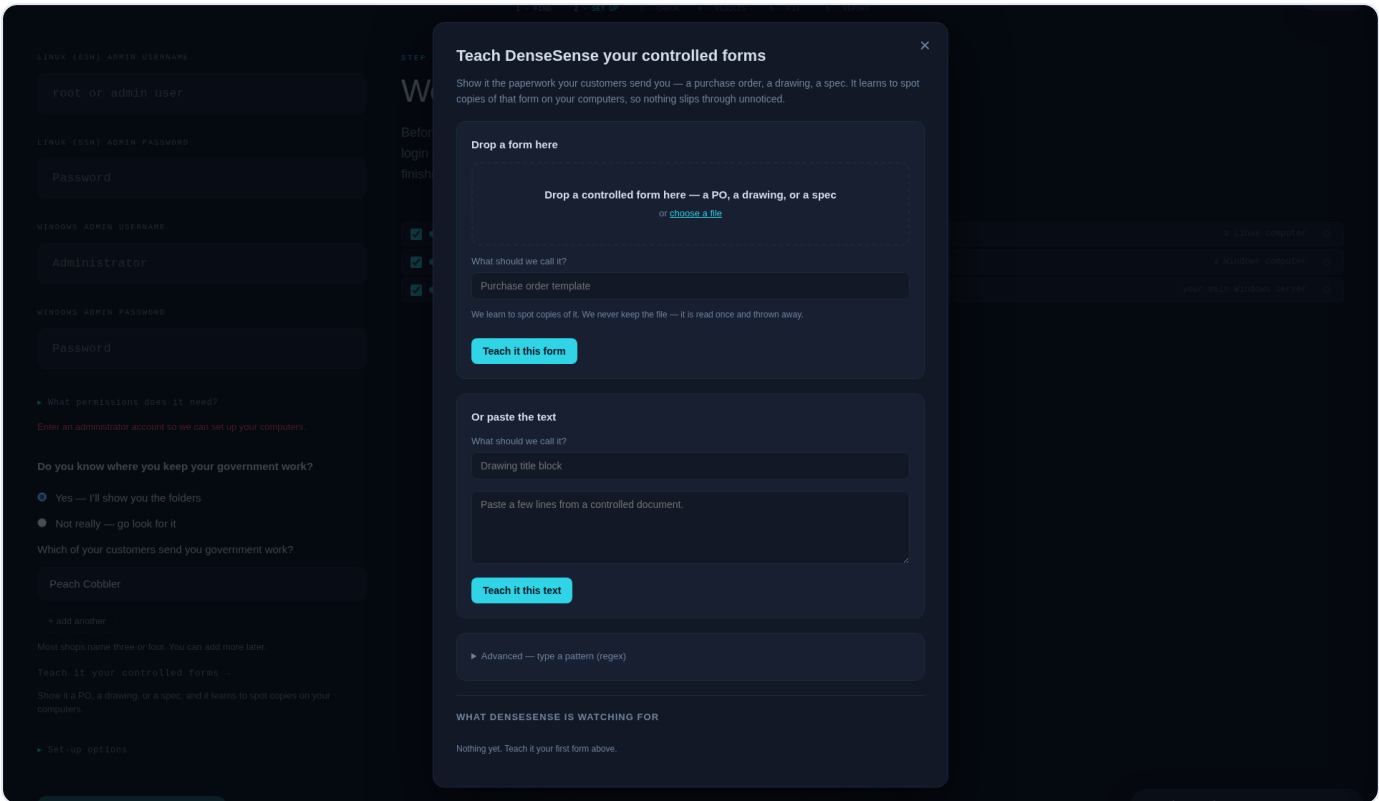
5 | What it reads -- and the one gap it names

Markings live in predictable places, and DenseSense reads exactly those places with the Python standard library -- **no new dependency ships to find them**. Office documents (`.docx` , `.xlsx` , `.pptx`) are ZIP archives of XML, and the banner and portion markings sit in the header and footer parts (`word/header1.xml` and friends), which a body-text reader misses entirely. DenseSense reads the parts that matter.

SOURCE	HOW
<code>.docx</code> / <code>.xlsx</code> / <code>.pptx</code>	Stdlib <code>zipfile</code> + XML -- headers and footers included, and runs re-joined so a marking split across <code><w:t>CUI//</w:t><w:t>SP-CTI</w:t></code> reads as the single token <code>CUI//SP-CTI</code>
PDF text	The text layer is read; a banner is required on every page, so the first pages are dispositive and a page cap costs nothing on a real document
Scanned images	Gap, stated: a marking printed inside a scanned image has no text layer and cannot be seen without OCR. DenseSense records the file as image-only rather than passing it as clean

6 | Teach it your controlled forms (the DECLARED tier)

Marking-detection is reliable and *blind exactly where the exposure lives*: a prime sends a purchase order, the contract says it is controlled, and the document carries no banner -- because primes mark inconsistently and the customer is not about to add one. Nothing in the text betrays it. The only source of truth for that document is **you**: you know because the government told you. DenseSense lets you teach it -- a document, pasted text, or a pattern -- and grades every resulting match as `DECLARED` so the provenance stays visible all the way to the report:



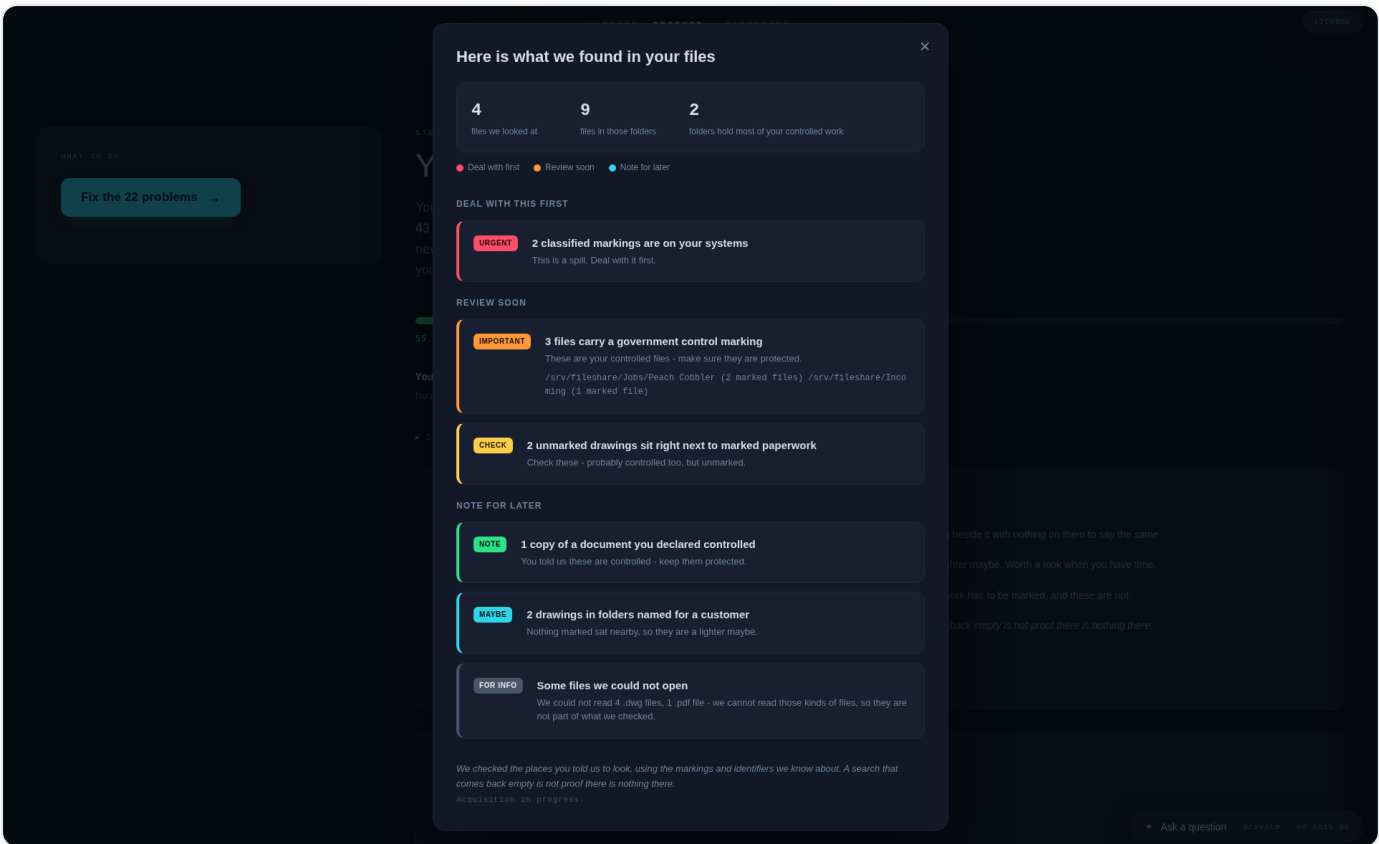
The teach modal: drop a controlled form, paste its text, or add a pattern. DenseSense derives a signature and discards the document; matches to it are graded DECLARED.

ENDPOINT	TEACH IT BY...
<code>POST</code> <code>/api/densesense/declare</code>	Pasting <code>{label, text}</code> or uploading <code>{label, file}</code> of a document you have designated controlled
<code>POST</code> <code>/api/densesense/regex</code>	A raw regex rule <code>{label, pattern}</code> for a form field or identifier you know

High-impact -- The document you teach is CUI by definition, so DenseSense **never keeps it**: an uploaded file is written to a temp path only long enough to extract from, then discarded; what persists is a derived signature, not your document. And a customer-supplied regex is a denial-of-service risk -- `cui_rules` safe-compiles it and statically refuses catastrophic-backtracking (ReDoS) shapes rather than letting one hang the scanner.

7 | What the report says -- container, not content

DenseSense reports at the level you can act on: the **container**. A finding names the file, the marking *type*, the tier, and a count -- and rolls up to *where the controlled work concentrates*, because "this share holds 14 CUI-marked files, here are the paths" is a decision and "3,182 matches" is noise. In the dashboard, that roll-up is the Step-4 intel modal:



The Step-4 intel modal: where the controlled work concentrates, graded by tier and location. Per-file paths and counts; no matched content is ever shown or stored.

OUTPUT	WHAT IT GIVES YOU
Per-file finding	Path, marking type, tier, count -- enough to open the file and confirm
Heatmap	Where findings concentrate, by directory -- the shares that actually hold the work
Concentration	The fewest directories that together hold most of the findings -- the short list to lock down first

High-impact -- No matched content, ever. A finding carries the marking type and a count -- never the matched text, the surrounding line, or an excerpt. Quoting content would place CUI inside a signed, retained evidence package handed to a third-party assessor, which would then inherit the handling requirements of the data it describes -- the exact exposure discovery is bought to prevent. The finding record has no field capable of holding content, and a test enforces it.

8 | The API surface

Everything the dashboard does, the API does too -- gated in one place (`_densesense_enabled()`); on a build where the capability is not enabled the modules are not even present, and the routes report it as unavailable rather than a half-working feature.

ENDPOINT	DOES
<code>POST</code> <code>/api/densesense/scan</code>	Start a walk over the roots you supply. Returns <code>202</code> and a <code>job_id</code> .
<code>GET</code> <code>/api/densesense/scan/<job_id></code>	Poll progress and read the tiered result when it completes.
<code>POST</code> <code>/api/densesense/scan/<job_id>/abort</code>	Stop a running walk.
<code>POST</code> <code>/api/densesense/declare</code>	Teach a controlled form -- text or file (Section 6). Hash-and-discard.
<code>POST</code> <code>/api/densesense/regex</code>	Teach a ReDoS-guarded regex rule.
<code>GET</code> <code>/api/densesense/rules</code> <code>DELETE</code> <code>/api/densesense/rules/<id></code>	List or remove the forms and rules you have taught.

9 | Where it fits in compliance

NIST SP 800-171 is about protecting CUI, and **you cannot scope, protect, or attest to what you have not located**. DenseSense sits in front of the assessment: it turns "we think the CUI is on the engineering share" into a finding with paths behind it. It is also a direct implementation of the data-inventory idea in CIS Controls v8.1 -- Safeguard 3.2 (data inventory) and 3.7 (data-classification scheme).

Tip -- Rev 2 of 800-171 (110 requirements) is the measurement baseline today. The Rev 3 restructure is pending as an Interim Final Rule (RIN 0790-AM01) and is **not in force** -- so DenseSense maps findings to controls at the reporting boundary, never by hard-coding a control ID into the detector, and this guide does not quote a Rev 3 date as if it had landed.

A clean scan is itself evidence: "we walked 412 GB across six hosts and found CUI markings in one share, here are the paths" is a stronger position in an assessor interview than a confident claim with nothing behind it.

10 | Early Adopter limitations & what is not built

Stated plainly, so nothing here is a surprise later:

- **Markings-focused only.** No PII, PHI, or cardholder-data detection. Those regimes need their own validated detectors (pattern + corroboration + checksum), which DenseSense does not ship.
- **No OCR.** A marking that exists only inside a scanned image is not read. The file is recorded as image-only rather than silently passed as clean.
- **Content-CUI is a lead, not a finding.** Unmarked CUI is found only when you teach it (the **DECLARED** tier). DenseSense will not guess a document into CUI on content alone and call it confirmed.
- **Early Adopter.** The capability is available to early adopters and usable, but it is early: expect the tier taxonomy and the operator surface to keep tightening.

Note -- These are the boundary of the Early-Adopter release, not a roadmap commitment. Treat any item here as "not available today" for planning purposes, and confirm current status against the build you are running.

DenseSense (Early Adopter) -- a standalone, separately-licensed DenseDefense product for regulated-data discovery that can also run as a ForteFide add-on; this guide covers driving it from ForteFide v26.09.06.0122, in active development. It runs on the one host you control and walks the roots you point it at; no cloud call, no telemetry, and no matched content ever leaves the finding. This guide describes shipped behavior and states its current limits plainly; it is not a commitment to a delivery date for the items under Section 10. [Contact us for early access.](#)

DenseDefense | densedefense.com. CUI is a designation applied by the U.S. Government or by contract; DenseSense locates evidence of that designation, it does not make the designation. Confirm scope decisions with your own compliance authority.