**Guideline & Installation**

**Module D – DeepShield Platform for Management & Integration**

DeepShield: A distributed and cryptographic approach for authenticating digital content, based on Trusted Execution environments (TEE), establishing ownership, and detecting unwanted manipulations with focus on the DeepShield Platform for Management & Integration with connection to the overall DeepShield approach.

Stage 1 (Months 1–7, timeboxed, final demonstration 27/28th Mail 2025):

Guideline how to setup and use the DeepShield Platform (Module D) demonstrated for phase Stage 1.

|  |  |
| --- | --- |
| Document No.: | **D\_Platform\_MI\_Guideline** |
| Issue: | A |
| Date: | 05.05.2025 |

|  |  |
| --- | --- |
|  |  |

**Revision Record**

|  |  |  |  |
| --- | --- | --- | --- |
| **Issue** | **Date** | **Description / Change** | **Modified Chapters** |
| A | 05.05.2025 | Delivered version | All |

Contact: langewisch@secublox.com

# **Table of contents**

[Table of contents 2](#_Toc197260086)

[Introduction 3](#_Toc197260087)

[Overall Goal DeepShield Approach 3](#_Toc197260088)

[Focus Module D 3](#_Toc197260089)

[List of Abbreviations 4](#_Toc197260090)

[DeepShield Overall Approach 5](#_Toc197260091)

[DeepShield Flow 6](#_Toc197260092)

[Setup integration scripts 12](#_Toc197260093)

[Integration 13](#_Toc197260094)

[Images 14](#_Toc197260095)

[Register authenticated Image 19](#_Toc197260096)

# Introduction

The rapid development of digital technologies in recent years has opened both fascinating and unsettling possibilities, particularly **in** synthetic content and deepfakes. This sophisticated AI-generated audio, image and video content is now so realistic that it is almost indistinguishable from real footage.

With the exponential increase in the performance of systems and language models based on artificial intelligence (AI) and machine learning (ML), not only the quantity but the quality of deepfakes is improving rapidly. This poses dangers – for trust, security and the perception of reality in our society.

Efforts to combat deepfakes are currently focused on two major areas:

* **Detection:** Development of AI algorithms to identify deepfakes
* **Prevention:** Implementation of authentication mechanisms for digital content.

Despite significant progress in these areas, major challenges remain, such as the generalizability and scalability of AI deepfake detection systems and the establishment of a manipulation-resistant standard for image metadata.

## Overall Goal DeepShield Approach

Development of a prototype that reliably detects deepfakes images from various media content and/or protects existing infrastructures from the use of deepfakes with preventive measures. To address these goals the overall DeepShield approach was divided between different independent modules (A.1, A.2, B, C, D) and will be integrated to complete the overall DeepShield approach, demonstrated at the end of stage 1.

## Focus Module D

This document focusses on the guideline, to setup and use module D a prototype demonstration of the DeepShield Platform integrated into the overall DeepShield approach.

A screenshot of a calendar

AI-generated content may be incorrect.

Figure 1: DeepShield – Stage 1 – D –DeepShield Platform

This document focuses on the setup and use of module D.

## List of Abbreviations

Below in Table 1 is a list of abbreviations used in this document.

|  |  |
| --- | --- |
| **Abbreviation** | **Description** |
| BC | Blockchain |
| DLT | Distributed Ledger Technology |
| DN | Decentralized Network |
| DSC | DeepShield |
| IPFS | InterPlanetary File System |
| KYC | Know Your Customer |
| LLM | Large Language Model |
| NFT | Non-Fungible Token |
| PRNG | Pseudo-Random Number |
| TEE | Trusted Execution Environments |
| TRL | Technology Readiness Level |
|  |  |
|  |  |
|  |  |

## DeepShield Overall Approach

The DeepShield approach will focus on following requirements to demonstrate the defined goals by following functionalities:

|  |  |
| --- | --- |
| **A: Content Creation & Secure Watermarking** |  |
| A user captures a photo or video with on a public blockchain registered device.  These photos or videos are watermarked on-device within a secure, trusted execution environment making use of the private device key and then stored locally or elsewhere. |  |
|  |  |
| **B: Content Registration**  Additionally, to the preregistered public key of the devices, the original watermarked content can be registered on a public blockchain. This allows for ensuring context integrity using further meta information such as reference vector embeddings. |  |
| **C: Content Verification & Context Integrity**  The registered public key enables verification of the content’s authenticity, confirming its origin from a specific device based on the watermark. Degradations in the watermark indicate content manipulation.  Using meta information such as reference vector embeddings, contextual integrity can be verified, in a multilayered approach, helping to prevent disinformation. |  |
| **D: Verification and Revocation via Distributed Ledger**  Public Key Management: Verifiers can retrieve the device’s public key from the blockchain to authenticate the watermark and digital signature, ensuring transparent, verifiable content authenticity.  Revocation Notices: The ledger will contain information about revoked or compromised keys, ensuring that only active and trusted devices can be verified  Smart Contracts for Automation: Implement smart contracts to handle device registration, key updates, and revocation seamlessly |  |

# DeepShield Flow

DeepShield flow to demonstrate to use Blockchain based registered digital identities (devices, accounts) to create authenticated images signed by registered digital identities and watermarked images.

**A diagram of a software development process

AI-generated content may be incorrect.**

Figure 1: DeepShield – Stage 1- Module A/B – Create and register images

First step will focus on create, embed watermarking and register images in an authenticated way, with focus on data centric security[[1]](#footnote-2). The overview flow is based on following view.

A screenshot of a computer

AI-generated content may be incorrect.

Figure 1: DeepShield – Stage 1- Module A/B – Create and register images - steps

|  |  |  |  |
| --- | --- | --- | --- |
| **Step** | **Test Data & Code-** **Snippets** | **Comment** | **Blockchain** |
| **0:DeepShield Smart-Contract** | DeepShield Contract issued by Secublox-Account  Secublox-DeepShield-Library:  # Smart-Contract instance of DeepShield  contract\_instance =  initialize\_connection(node\_url, smart\_contract\_address, abi\_file)  # example python script  https://gitlab.com/secublox-platform/iabg\_deepshield/-/blob/main/Module\_B\_D/images.py | Smart-Contract to register, manage & execute transactions | <https://sepolia.arbiscan.io/address/0xBFd555e7BF08e6EEa58Aa2120a0265136d952Ca5> |
| **0:user account creator (user) registered** | Account: iabg\_creator\_1  Wallet (public key):  0xBA4d9B240F88856E70D25AAce76F99878d1cC947  Wallet (private key):  0xf21d1e7f5201c008a87454fba2d80c6016bfdb3da343145e1a8fdf7fd5757d37 | User accounts are able to register and assign devices / apps | <https://sepolia.arbiscan.io/address/0x1dE101999f77fDa9941A9E440210d819ED7F33BC> |
| **0:app account creator (device/app) registered** | Account: iabg\_cc\_app\_1  Wallet (public key):  0x6841E3AEFB55Eb087F923fd2971Bc642b858F79b  Wallet (private key):  0xaba03fd5f886d97052fb26be51847e2f7de3e7d86b34f234ee2ec797e114a113 | app account is able to register images | <https://sepolia.arbiscan.io/address/0x6841E3AEFB55Eb087F923fd2971Bc642b858F79b> |

|  |  |  |  |
| --- | --- | --- | --- |
| Create, watermark and register image (executed inside TEE environment) | | | |
| **1. Create image** | By:  User Account: iabg\_creator\_1  App Account: iabg\_cc\_app\_1  Type: JPEG, e.g. image-RAW | Image created and locally (e.g. file system) stored |  |
| **1.1 Add image** | DeepShield-UI:  <https://deepshield.secublox.com/images>  A screenshot of a computer  AI-generated content may be incorrect. | Add image without encryption and assigned with content creator: iabg\_CC\_app\_1 | Added image raw data available / stored in decentralized network (->UI: image view)A person in a suit  AI-generated content may be incorrect.  Image raw data:  A person in a suit  AI-generated content may be incorrect. |

|  |  |  |  |
| --- | --- | --- | --- |
| **2. Embade watermark with image** | DeepShield-UI:  A screenshot of a computer  AI-generated content may be incorrect.  Select value for “alpha”  IABG-Implementation:  # Embed the watermark into the image using the selected method  watermarked\_image, ground\_truth\_watermark = watermarking\_method.embed(  image=original\_image,  watermark=watermark,  watermark\_positions=watermark\_positions,  alpha=alpha,  ) | Configure “alpha” to define intensity / visibility of the watermark | Status about watermarking stored in decentralized network (->UI: image view)  A screenshot of a computer  AI-generated content may be incorrect.  Watermarked image:  A person in a suit  AI-generated content may be incorrect. |

|  |  |  |
| --- | --- | --- |
| **3. Encrypt image (optional)** | DeepShield-UI:  A screenshot of a computer  AI-generated content may be incorrect. | A screenshot of a computer  AI-generated content may be incorrect. |
|  | Encrypred matermarked image:  A screenshot of a computer  AI-generated content may be incorrect. | |

|  |  |  |
| --- | --- | --- |
| **4 Register image and ownership** | Ownership token of the image  minted for wallet referenced with image content.  Image raw data registered  Data matrix registered | Example :  <https://sepolia.arbiscan.io/tx/0x8f14c1a6eec5c0b23dff9a1ddf909cc07b04b10c1e79860b47843ed047f357d9>  A screenshot of a computer  AI-generated content may be incorrect. |
|  | A screen shot of a computer  AI-generated content may be incorrect.  # example python script  <https://gitlab.com/secublox-platform/iabg_deepshield/-/blob/main/Module_B_D/images.py> | |

# Setup integration scripts

To connect with Blockchain network and register images following activities are needed upfront. This software library (based on Python) connects account(creator, verifier), content creator(like a device or app), content verifier (like ai, deepfake) to the blockchain network in order to initialize, read and write blockchain transactions.

*# update local environment*

sudo apt update

*#* ***Python Version 3.12.3***

*# install python3*

sudo apt install python3-pip

*# install Secublox python package:*

python3 -m pip install --index-url <https://test.pypi.org/simple/> --no-deps package-deepshield-lib==0.0.4

# install web3 package library

pip install web3==6.0.0

# install tools (MacOs)

pip install setuptools

DeepShield-Smart-Contract (e.g.

address: 0x7830f9e05beB24D997b0534590fa71a623B50952

explorer: <https://sepolia.arbiscan.io/address/0xBFd555e7BF08e6EEa58Aa2120a0265136d952Ca5>

) will be updated in parallel while development and to use this library it needs the newest version together with the ABI-file (contract.abi).

# Integration

The dashboard for DeepShield-Platform is available via Secublox cloud services by following URL: <https://deepshield.secublox.com>

A screenshot of a login form

AI-generated content may be incorrect.

Figure 3: DeepShield-Platform Login

Access to the DeepShield-Platform is available by following IABG user-accounts integrate and test the logic.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Account** | **Password** | **Profile** | **Comment** | **Wallet (public key)** |
| iabg\_creator\_1 | @DEEPSCAM2025! | creator | User account with access to DeepShield (Platform, API, Smart-Contract) to manage registered images and assigned content creator app | 0xBA4d9B240F88856E70D25AAce76F99878d1cC947 |
| iabg\_cc\_app\_1 | @DEEPSCAM2025! | App | Raspberry PI / Smartphone (Android) to register authenticated / watermarked images, registered by iabg\_creator\_1 | 0x6841E3AEFB55Eb087F923fd2971Bc642b858F79b |
| iabg\_verifier\_1 | @DEEPSCAM2025! | verifier | Access to DeepShield (Platform, API, Smart-Contract) to verify created images and verification services | 0x72bb13E5d26F4D5EaEDb14E6a2d5CCED62Cd3677 |
| iabg\_check\_1 | @DEEPSCAM2025! | ai | Now access all the images and able to verify the image by verification services | 0x628F07ed04f1a420d7eB4c6223f2EcE685bcFcD7 |

## Images

After login with the account “iabg\_cc\_app\_1” it is possible, to see an overview of all by this device registered images in the Blockchain.

A screenshot of a computer

AI-generated content may be incorrect.

Figure 4: DeepShield – Images table view

After open the info of a single image a popup-window shows the image view together with metadata.

The DeepShield platform supports 3 types of image encryption.

1. **No encryption image** when we open image-info popup.

|  |  |
| --- | --- |
|  |  |
| Figure 5: DeepShield – Images view | Figure 6: DeepShield – Images properties view |

After click on meta link the platform redirects to new window with assigned meta information’s.

A screenshot of a computer

AI-generated content may be incorrect.

Figure 7: DeepShield – Image meta info

eg. link :- <https://deepshield.mypinata.cloud/ipfs/QmX6D92yUM7Vaj8JvvLHzfqYFiVGsC4F5bWRX8QWpSnX4Q?pinataGatewayToken=cJi7ejhGVyFet2kGKZLUPEWnOnqHe8KHbgN5z8KP3bEoOQEw6nQ-mEsQAELK6AWd>

When click on Content Creator (device/app) or Content Creator (Account) then we just redirected to wallet with all the transaction

A screenshot of a computer

AI-generated content may be incorrect. A screenshot of a computer

AI-generated content may be incorrect.

Figure 8: DeepShield – Image wallet info

1. **Encryption** **with private key** image when we open image-info popup.

|  |  |
| --- | --- |
|  |  |
| Figure 9: DeepShield – Images view | Figure 10: DeepShield – Images properties view |

After click on meta link the platform redirectes to new window with the meta data information’s.



Figure 11: DeepShield – Image meta info

eg. link :- <https://deepshield.mypinata.cloud/ipfs/QmYn68nhK19hyP1cSWjpkNTep4jQaYQfzyFtfxJ2p9aGXu?pinataGatewayToken=cJi7ejhGVyFet2kGKZLUPEWnOnqHe8KHbgN5z8KP3bEoOQEw6nQ-mEsQAELK6AWd>

When click on **Content Creator (device/app) or Content Creator (Account)** then the platform redirectes to the Blockchain wallet with all underlying Blockchain transactions.

A screenshot of a computer

AI-generated content may be incorrect. A screenshot of a computer

AI-generated content may be incorrect.

Figure 12: DeepShield – Image wallet info

1. **Encryption** **with AES 256 key** image when we open image info popup.

|  |  |
| --- | --- |
| A screenshot of a computer  AI-generated content may be incorrect. | A screenshot of a computer  AI-generated content may be incorrect. |
| Figure 13: DeepShield – Images view | Figure 14: DeepShield – Images properties view |

After click on meta link the platform redirectes to new window with meta information’s.

A black screen with white lines

AI-generated content may be incorrect.

Figure 15: DeepShield – Image meta info

eg. link :- <https://deepshield.mypinata.cloud/ipfs/QmZumGEqkAW8ZN3gcesDk43eyVsBxBtjpgsyaH7yj4xLpY?pinataGatewayToken=cJi7ejhGVyFet2kGKZLUPEWnOnqHe8KHbgN5z8KP3bEoOQEw6nQ-mEsQAELK6AWd>

After click on **Content Creator (device/app) or Content Creator (Account)** then it redirectes to the underlying wallet with all assigned Blockchain transactions.

A screenshot of a computer

AI-generated content may be incorrect. A screenshot of a computer

AI-generated content may be incorrect.

Figure 16: DeepShield – Image wallet info

If the image is registered and when we click on register of **DeepShield** column then the platform redirects to Blockchain transaction details.

A screenshot of a computer

AI-generated content may be incorrect.

Figure 17: DeepShield – Image wallet info

## Register authenticated Image

After login with the account “iabg\_creator\_1” it is possible, to add new image.

1. **Encryption Type => No Encryption**

A screenshot of a computer

AI-generated content may be incorrect.

Figure 18: DeepShield – Add new image

1. **Encryption Type => Private Key**

A screenshot of a computer

AI-generated content may be incorrect.

Figure 19: DeepShield – Add new image

1. **Encryption Type => AES 256 Key**

After select option of encryption with AES 256 key, then a select AES key from drop-down is together with a content creator (device/app) is needed.

A screenshot of a computer

AI-generated content may be incorrect. A screenshot of a computer

AI-generated content may be incorrect.

Figure 20: DeepShield – Add new image

After login with the account “iabg\_cc\_app\_1” it is possible to register the created image.

A screenshot of a computer

AI-generated content may be incorrect.

Figure 21: DeepShield – Register image

**A screenshot of a computer

AI-generated content may be incorrect.**

Figure 22: DeepShield – Register image

1. https://www.paloaltonetworks.com/cyberpedia/data-centric-security [↑](#footnote-ref-2)