The Printer Working Group

Imaging Device Security

August 10, 2023

PWG August 2023 Virtual Face-to-Face

Agenda

Please Note: This PWG IDS Meeting is Being Recorded

When	What
10:00 – 10:05	Introductions, Agenda review
10:05 – 10:45	Discuss status of HCD iTC, HIT and plans for future HCD cPP/HCD SD releases
10:45 – 11:25	AI Cybersecurity in the EU and US
11:25 – 11:30	HCD Security Guidelines v1.0 Status
11:30 – 11:55	TCG/IETF Liaison Reports
11:55 – 12:00	Wrap Up / Next Steps

Antitrust and Intellectual Property Policies

"This meeting is conducted under the rules of the PWG Antitrust, IP and Patent policies".

- Refer to the Antitrust, IP and Patent statements in the plenary slides

Officers

- Chair:
  - Alan Sukert
- Vice-Chair:
  - TBD
- Secretary:
  - Alan Sukert
- Document Editor:
  - Ira McDonald (High North) – HCD Security Guidelines

HCD international Technical Community (iTC) Status

HCD international Technical Community (iTC) Status

- Since last IDS F2F on May 18, 2023 HCD iTC meetings have been held on:
  - June 26th
  - July 17th

  NOTE: Since publishing the HCD cPP v1.0 and HCD SD v1.0 in Oct 2022 the HCD iTC has gone to meeting once a month

- Current focus is on:
  - Developing a release plan for future versions of the HCD cPP and HCD SD
  - Determining content for and then implementing the next HCD cPP / HCD SD release
  - Addressing issues against HCD cPP / SD v1.0

HCD cPP/SD v1.0 Status

- Version 1.0 of both documents published on October 31, 2022
- Awaiting Endorsements from NIAP (US), ITSCC (Korea), JISEC (Japan)
  - NIAP and the Canadian Scheme are currently reviewing the HCD cPP (see HIT Slide)
  - CCDB is reviewing the HCD SD
  - Other Schemes (not sure which ones) are reviewing the HCD cPP
  - As of now have no status on ITSCC or JISEC
  - May get some Endorsements at the Fall CCDB Meetings in Wash DC
- Canadian Scheme issued an Endorsement in Feb 2023
  - A vendor (Lexmark) is actively pursuing certification of an HCD against HCD cPP / HCD SD v1.0

HCD cPP/SD
HCD Interpretation Team (HIT) Status

- HIT now has 10 members
  - Current HIT membership consists of HCD vendors (5), Evaluation Labs (2), Consultant (1) and Schemes (NIAP and Canadian)
  - Meets desired maximum of 10 members on the HIT
- HIT procedures v1.0 now finalized and infrastructure set up
  - Using GitHub for documenting Requests for Interpretation (RfIs) and for creating and tracking changes to HCD cPP v1.0 and HCD SD v1.0 for approved RFIs
  - Created new HCD-IT repository and Integration baseline for changes approved by the HIT
- Have had six HIT Meetings so far to review and process issues submitted for RfIs and approve HIT procedures v1.0 – See next 8 slides

HCD cPP/SD HIT RfI Status

Issue #	Title	Issue	Status
HCD-IT #1	The FCS_COP.1/KeyEnc Cryptographic operation (Key Encryption) SFR in HCD cPP v1.0 is inconsistent with TPM 2.0 Architecture specification section 26.6 “Sensitive Area Encryption"	FCS_COP.1/KeyEnc SFR - Case: AES algorithm • AES used in [[selection: CBC, GCM] mode] TPM 2.0 Architecture specification Section 26.6 (Page 172) - "All symmetric encryption of the sensitive area uses Cipher Feedback (CFB) mode." CFB is the only AES mode allowed by the TPM 2.0 specification	Under Review – Looking at alternatives approach of using FPT_KYP_EXT.1.1 Key Protection SFR option. **No change since last IDS Session**
HCD-IT #2	Clarification is needed about algorithm verification of Root of Trust in the Test Assurance activities for the Secure Boot SFR	HCD SD Section 2.6.1 FPT_SBT_EXT.1 Extended: Secure Boot, 2.6.1.3 Tests, pg. 59: Add a note in this section saying that the algorithm verification for Root of Trust should be avoided, because authenticity check in Root of Trust should be performed by some kind of immutable code, so the algorithm verification tests should be difficult to perform.	In Progress – Solution has been developed; Technical Decision being prepared.

HCD cPP/SD HIT RfI Status

Issue #	Title	Issue	Status
HCD-IT #3	Extraneous "selection" in SFR FCS_CKM.4 Cryptographic key destruction in HCD cPP v1.0	Section 5.3.5, FCS_CKM.4 Cryptographic key destruction on page 33: in FCS_CKM.4.1 the last line of the SFR states "] that meets the following: [selection: no standard]." Since the selection has already been made in the cPP, the "selection:" should be deleted.	Complete - Issue was closed with no action taken since it was a duplicate to one of the samples indicated in Issue HCD-IT #7
HCD-IT #4	NIAP APE_ECD.1-5 Evaluation Comments against the HCD cPP	As part of NIAP’s review process of the HCD cPP, we performed an evaluation of the APE work units and identified several needing correction. Please see the following comments: APE_ECD.1-5, The evaluator shall examine the extended components definition to determine that each extended functional component uses the existing CC Part 2 components as a model for presentation. – Gave several example	Awaiting Review – HCD-IT #4- #7 are part of the NIAP evaluation of the HCD cPP as part of the certification of the HCD cPP. All the examples and general comments provided by NIAP must be fixed and included in an update to v1.0 as quickly as possible **No change since last IDS Session**

HCD cPP/SD
HIT RfI Status

Issue #	Title	Issue	Status
HCD-IT #5	NIAP APE_REQ.2-5 Evaluation Comments against the HCD cPP	As part of NIAP's review process of the HCD cPP, we performed an evaluation of the APE work units and identified several needing correction. Please see the following comments: APE_REQ.2-5, The evaluator shall examine the statement of security requirements to determine that all assignment operations are performed correctly. – provides several examples	See HCD-IT #4 **No change since last IDS Session**
HCD-IT #6	NIAP APE_REQ.2-8 Assessment Comments against the HCD cPP	As part of NIAP's review process of the HCD cPP, we performed an evaluation of the APE work units and identified several needing correction. Please see the following comments: APE_REQ.2-8, The evaluator shall examine the statement of security requirements to determine that all refinement operations are performed correctly. -- general inconsistency as to whether an SFR with a refinement in it starts with "Refinement:" or not – several examples noted	See HCD-IT #4 **No change since last IDS Session**

HCD cPP/SD
HIT RfI Status

Issue #HCD	Title	Issue	Status
HCD-IT #7	NIAP APE_REQ.2-7 Assessment of HCD cPP	As part of NIAP's review process of the HCD cPP, we performed an evaluation of the APE work units and identified several needing correction. Please see the following comments: APE_REQ.2-7, The evaluator shall examine the statement of security requirements to determine that all selection operations are performed correctly. -- General inconsistency with regards to whether or not "selection:" prompt is bolded Examples are provided	See HCD-IT #4 Will also include two similar comments from the Canadian Scheme **No change since last IDS Session**

HCD cPP/SD
HIT RfI Status

Issue #HCD	Title	Issue	Status
HCD-IT #8	Update of Application Notes in SFR FPT_KYP_EXT.1 Needed in HCD cPP v1.0 to Clarify Key Storage Conditions	In the discussions by the HIT of Issue HCD-IT #1, one proposed solution to the issue was to use the provisions of SFR FPT_KYP_EXT.1 Extended: Protection of Key and Key Material to address the concern expressed in HCD-IT #1. However, during the discussion it was pointed out that one deficiency of the FPT_KYP_EXT.1 in HCD cPP v1.0 is that the Application Notes for this SFR d not adequately explain what all the conditions in SFR FPT_KYP_EXT.1.1 that pertain to the storage of keys are. This issue is to request that the Application Notes in SFR FPT_KYP_EXT.1 be modified to more clearly explain what each of the conditions for key storage mean in SFR FPT_KYP_EXT.1.1.	Awaiting Review – Working on the exact wording of the revised Application Notes for SFR FPT_KYP_EXT.1. **New since last IDS Session**

HCD cPP/SD
HIT RfI Status

Issue #HCD	Title	Issue	Status
HCD-IT #9	Modification proposal : tests for FDP_DSK_EXT.1.	This SFR should be satisfied and certified if encryption of any confidential data will not depend on a user electing to protect that data. However current test description is limited to perform writing to the storage device with "operating TSFI" which enforce write process of User documents and Confidential TSF data. Therefore, a functionality which does not have such TSFI and the data cannot be tested and certified even if the TOE function is satisfied with the SFR. This situation should be corrected. For more detail, SWAP and Core dump etc., are written User documents and Confidential TSF data to storage device by system (OS) at any timing as necessary. SWAP and Core dump etc., doesn't write any User documents and Confidential TSF data when TSFI is operated.	Awaiting Review – this was a legacy issue. The issue is in Section 3.1.3.4 of the SD; specifically Test 1 for FDP_DSK_EXT.1which explicitly requires an operating TSFI, but encryption of data stored on a storage device needs to be done without user intervention, meaning there is no TSFI involved. Working to modify the tests in Section 3.1.3.4 to remove the references to TSFIs in verifying the written data is properly encrypted **New since last IDS Session**

HCD cPP/SD
HIT RfI Status

Issue #HCD	Title	Issue	Status
HCD-IT #10	Mapping issue between Mandatory 'O.KEY_MATERIAL' objective and Cond. Mandatory FPT_KYP_EXT.1 **NOTE: New since last IDS Session**	[APE_REQ.2-11] According to HCD cPP I.6, "O.KEY_MATERIAL" is defined as a mandatory objective. I.9 maps "O.KEY_MATERIAL" only to "FPT_KYP_EXT.1" which is a conditionally mandatory SFR. This creates scenarios where mandatory "O.KEY_MATERIAL" security objective cannot be satisfied when FPT_KYP_EXT.1 is not claimed as per conditions are not met (Section 1.4.2 "USE CASE 2: Conditionally Mandatory Use Cases"). Additional details: I reviewed the following GitHub issue on cPP Draft where I believe the decision was made to remove O.KEY_MATERIAL from being "conditionally mandatory". However, no information is found on how this change affects the mapping: HCD-iTC/HCD-iTC-Template#238	Awaiting Review – This came rom the Canadian Scheme's review of HCD cPP v1.0. Issue is that OSP O.KEY_MATERIAL is mapped to SFR **FPT_KYP_EXT.1** which is a "Conditionally Mandatory" SFR. Means that OSP O.KEY_MATERIAL would only apply conditionally in cases where , for example, an HCD had hard disks and would not apply to TSF data stored in wear-leveling devices such as SSDs. Best solution is to map O.KEY_MATERIAL to a mandatory SFR like **FPT_SKP_EXT.1 Extended: Protection of TSF Data**.

HCD cPP/SD
HIT RfI Status

Issue #HCD	Title	Issue	Status
HCD-IT #11	In FCS_CKM.4 Cryptographic key destruction, clarification needed whether encrypted keys stored in non-volatile memory are within the scope of key destruction **NOTE: New since last IDS Session**	This issue was submitted by Shin-ichi Inoue of Ecsec Laboratory For Section 5.3.5 **FCS_CKM.4 Cryptographic key destruction** in the HCD cPP, it is not clear that encrypted keys stored in non-volatile memory is within the scope of key destruction. Suggested change is to describe in an Application Note whether encrypted keys stored in non-volatile memory are within the scope of key destruction or not.	Awaiting Review –The key to the issue is the word "encrypted" in the Issue statement. This Issue is also linked to Section 5.3.4, SFR FCS_CKM_EXT.4 Extended: Cryptographic Key Material Destruction and SFR 5.3.4.1 which states "FCS_CKM_EXT.4.1 The TSF shall destroy all plaintext secret and private cryptographic keys and cryptographic critical security parameters when no longer needed." The central question of this issue – should we be destroying all keys or just plaintext keys. The HIT is divided on the answer to this question.

HCD iTC
HIT Release Plan

- Will definitely need an Errata release ASAP to address, as a minimum, the comments from the NIAP and Canadian Schemes
  - May include fixes for one or more of the open issues (at the time of release) against HCD cPP and HCD SD v1.0
- There may be additional standalone HCD cPP or HCD SD v1.0.x releases after the initial Errata release. If so and how many of these releases will occur likely depend on the comments we get from:
  - The review of the HCD SD from the CCDB
  - The review of the HCD cPP from the other Schemes and
  - The current Lexmark certification and future certifications against HCD cPP or HCD SD v1.0 from the applicable Evaluation Lab or applicable Scheme

  Note: The nature and severity of the comments will probably determine whether comments against HCD cPP or HCD SD v1.0 get fixed in a v1.0 release or get fixed in a later version.

HCD iTC
Issues Post-Version 1.0 – Release Plan

- In the past release plans have been based on whether to have major releases on maybe a 2-3 year bases and minor releases on possibly 12 - 15 month basis as needed
- Now, several factors have forced release plans to be based on these four major factors that will help govern the future content on the HCD cPP and SD and the timing of that content:
  - CCDB Specification of Functional Requirements for Cryptography
  - CC:2022 Compliance
  - Syncing with ND cPP / SD v3.0
  - CNSA 2.0

HCD iTC
Issues Post-Version 1.0
CCDB Specification of Functional Requirements for Cryptography (aka the "Crypto Spec")

- Draft Specification from the Common Criteria Development Board (CCDB) Crypto Working Group of key cryptographic SFRs that are commonly used in cPPs
- Appears from examination of the draft document that the text of the SFRs in the draft Specification either:
  - Came from the CC:2022 FCS Class SFRs, although interestingly some of them were changed;
  - Were created by the CCDB Crypto Working Group; or
  - May have come from the modified text of crypto SFRs in various cPPs
- Review comments were due by July 31st. Some of the comments from the HCD iTC against the draft Specification were:
  - Are all of the SFRs in the Specification mandatory or can an iTC pick the ones they need like we can do in CC:2022 Part 2
  - Is "Exact Conformance" to the SFRs in the Specification required or can an iTC add additional requirements to the SFRs
  - If a PP or cPP already has a version of an SFR that is in the Specification that is different from the version in the Specification, are the iTCs required to use the version in the specification
  - Will the necessary Assurance Activities for each of the SFRs in the Specification be provided
  - What is the transition plan to SFRs in the Specification when published

HCD iTC
Some Key Differences in the Crypto SFRs Between the "Crypto Spec" and the HCD cPP

- Many SFRs in the Crypto Spec added additional algorithms, key sizes and applicable standards not included in the HCD cPP versions of those SFRs
- The Crypto Spec uses the FCS_CKM key management SFRs from CC:2022 which are different from the FCS_CKM key management SFRs in the HCD cPP. However:
  - Crypto Spec added two new key management SFRs - **FCS_CKM_EXT.7 Cryptographic Key Agreement** and **FCS_CKM_EXT.8 Password-Based Key Derivation** – that are not in CC:2022
  - Crypto Spec SFR made changes to the version of **FCS_CKM_EXT.3 Cryptographic Key Access** from CC:2022
- The Crypto Spec took the **FCS_RBG** family from CC:2022, but changed SFR **FCS_RBG.1.1** from the version of that SFR in CC:2022
- SFR **FCS_COP.1/SigGen Cryptographic Operation (Signature Generation and Verification** in the HCD cPP covered both Signal Generation and Signal Verification; the Crypto Spec has separate SFRs for Signal Generation (**FCS_COP.1/SigGen Cryptographic Operation (Signature Generation)**) and Signal Verification (**FCS_COP.1/SigVer Cryptographic Operation (Signature Verification)**)
- The Crypto Spec version of SFR **FCS_KYC_EXT.1 Extended: Key Chaining** is completely different from the version of this SFR in the HCD cPP

HCD iTC
Issues Post-Version 1.0
Transitioning to CC:2022

- CC v3.1 R5 is the last revision of version 3.1 and may optionally be used for evaluations of Products and Protection Profiles starting no later than the 30th of June 2024
- Security Targets conformant to CC:2022 and based on Protection Profiles certified according to CC v3.1 will be accepted up to the 31st of December 2027
- After 30th of June 2024, re-evaluations and re-assessments based on CC v3.1 evaluations can be started for up to 2 years from the initial certification date
- New initial certifications based on CC v3.1 R5 may be started until 30th of June 2024
  - **Product certifications based on CC v3.1 R5 against a PP or PP configuration claiming exact conformance may be started until 31st of December 2025**
  - PP authors must update the PP or PP configuration to CC:2022 as soon as possible, and any new or updated PPs or PP configurations published after 30th of June 2024 must be based on CC:2022
- After 30th of June 2024, re-evaluations and re-assessments based on CC v3.1 evaluations can be started for up to 2 years from the initial certification date

HCD iTC
Some Key CC:2022 Changes

- **FAU_GEN.1 Audit data generation** and some of the other FAU Class SFRs changed from requiring "audit reports" to requiring "audit data"
- **FCS_CKM.4 Cryptographic key destruction** was deprecated and replaced by a new SFR **FCS_CKM.6 Timing and event of cryptographic key destruction**
- In the SFR **FPT_STM.1 Time stamps**, a new SFR **FPT_STM.2.1 The TSF shall allow the [assignment: *user authorized by security policy*] to [assignment: *set the time, configure another time source*]].** was added
- Key new SFRs added:
  - **FAU_STG.1 Audit data storage location**
  - **FCS_CKM.5 Cryptographic key derivation**
  - **FCS_RBG.1 Random bit generation**

    Note: there is a set of five other FCS_RBG SFRs in CC:2022 that provide additional requirements beyond basic Random Bit Generation
  - **FCS_RNG.1 Random number generation**
  - **FDP_SDC.1 Stored data confidentiality**
  - **FIA_API.1 Authentication proof of identity**
  - **FTP_PRO.1 Trusted channel protocol**
  - **FTP_PRO.2 Trusted channel establishment**
  - **FTP_PRO.3 Trusted channel data protection**

HCD iTC
Issues Post-Version 1.0
ND cPP / SD v3.0 Content

Changes in ND cPP / SD v3.0 that could necessitate updates to existing SFRs / Assurance Activities or inclusion of new SFRs / Assurance Activities in updates to HCD cPP / SD:

- Claim conformance to NIAP Functional Package for SSH
- Updates to TLS and DTLS SFRs to incorporate TLS 1.3 and removal of TLS 1.1
- Inclusion of new SFRs - **FAU_STG_EXT.1 External Audit Trail Storage, FCS_TLSC_EXT.1 TLS Client Protocol Without Mutual Authentication, FCS_TLSS_EXT.1 TLS Server Protocol without Mutual Authentication, FCS_TLSS_EXT.2 TLS Server Support for Mutual Authentication, FCS_DTLSC_EXT.2 DTLS Client Support for Mutual Authentication** and **FPT_STM.1 Reliable Time Stamps (FPT_STM.1.2)**
- Inclusion of new SFRs **FCS_TLSC_EXT.3 TLS Client Support for secure renegotiation (TLSv1.2 only)** and **FCS_TLSS_EXT.3 TLS Server Support for secure renegotiation**
- Inclusion of Optional Security Assurance Requirements for Flaw Remediation (ALC_FLR)
- Added additional requirements to several crypto SFRs like FCS_CKM.4 Cryptographic Key Destruction and FCS_RBG_EXT.1 Random Bit Generation

Commercial National Security Algorithm (CNSA) Suite 2.0 Algorithms

Algorithm	**Function**	**Specification**	**Parameters**
Advanced Encryption Standard (AES)	Symmetric block cipher for information protection	FIPS PUB 197	Use 256-bit keys for all classification levels
CRYSTALS-Kyber	Asymmetric algorithm for key establishment	TBD	Use Level V parameters for all classification levels
CRYSTALS-Dilithium	Asymmetric algorithm for digital signatures	TBD	Use Level V parameters for all classification levels
Secure Hash Algorithm (SHA)	Algorithm for computing a condensed representation of information	FIPS PUB 180-4	Use SHA-384 or SHA-512 for all classification levels
Leighton-Micali Signature (LMS)	Asymmetric algorithm for digitally signing firmware and software	NIST SP 800-208	All parameters approved for all classification levels SHA256/192 recommended
Xtended Merkle Signature Scheme (XMSS)	Asymmetric algorithm for digitally signing firmware and software	NIST SP 800-208	All parameters approved for all classification levels

Original Detailed NIAP Transition Plan for CNSA Suite 2.0

- Currently all NIAP PPs must have CNSA 1.0 algorithms
- Will add SHA-512 to all NIAP PPs
- Will require either CNSA 1.0 or CNSA 2.0 be mandatory on all NIAP PPs
- Will implement CNSA asymmetric algorithms for software/firmware signing per following
  - LMS – 1H 2023
  - XMSS – 2H 2023
- Will implement following Key Establishment CNSA 2.0 algorithms in all NIAP PPs when they are standardized and all relevant Assurance Activities have been defined and agreed upon:
  - CRYSTALS - Kyber
  - CRYSTALS – Dilithium (used for Digital Signatures)
- Will deprecate CNSA 1.0 in 2030 – 2033 timeframe
- No current timeline established to make CNSA 2.0 mandatory
  - Will make use of CNSA 2.0 mandatory to be listed on PCL at some point
- Will work with vendors to help try to meet NSA schedule
- Will discuss with CCRA and engage with iTCs how best to integrate CNSA 2.0 into cPPs

HCD cPP/SD Content Post-Version 1.0 Potential V1.1 Content

- Incorporation of the SFRs from the CCDB Specification of Functional Requirements for Cryptography once it is published and we get a transition plan
  - We don�much know what either the CCDB or the various Schemes are going to require with respect to the "Crypto Spec" yet
- Updates for the relevant changes in CC:2022
- Inclusion of support for TLS 1.3 and deprecation of TLS 1.1, including updates to TLS and DTLS and other relevant changes per ND cPP/SD 3.0
- Incorporate the NIAP Functional Package for SSH so can claim conformance to it
- Inclusion of AVA_VAN and ALC_FLR.*
- Initial implementation of CNSA 2.0 algorithms
  - Inclusion of SHA-384 and SHA-512 and possible inclusion of LMS as an option likely first steps
- Changes due to any approved RfIs (Issues) to HCD cPP/SD v1.0
  - Will have to decide if only include changes approved by NIAP
- Inclusion of NTP
- Changes due to requests from JISEC, ITSCC, NIAP, Canada and possible other Schemes

HCD cPP/SD Content Post-Version 1.0 Potential for Inclusion in Later Versions

- **Full implementation of CNSA 2.0**
- **Support for any new crypto algorithms**
- **NIAP IPsec Package or other new NIAP Packages**
- **Updates due to changes from other ISO, FIPS or NIST Standards/Guidelines, and NIAP TDs**
- **Updates to Address 3D printing and the Digital Thread to Additive Manufacturing**
- **Support for Cloud Printing**
- **Support for Artificial Intelligence**
- **Support for Wi-Fi**
- **Any new CCDB Crypto WG or CCUF Crypto WG Packages or Specifications**
- Support for Security Information and Event Monitoring (SIEM) and related systems
- Support for SNMPv3
- Support for NFC
- Updates based on new technologies, customer requests or government mandates
- Syncing with newer updates to ND and FDE cPPs/SDs

HCD iTC Status
Key Next Steps

- Continue HIT activities for maintaining HCD cPP/SD v1.0 and issue the necessary TDs/TRs and Errata to address all documented RfIs
- Determine HCD cPP/HCD SD release plan for both v1.0 and updated versions
- Determine the content for and then create the next HCD cPP/SD releases for both v1.0 and v1.1 or V2.0, whichever is next
- Fully engage the HCD iTC to work on the next update to the HCD cPP and HCD SD
- Engage in long-range planning to determine what content will be needed in the HCD cPP/SD in the 3-5 year range and beyond

HCD iTC Status
Some General Lessons Learned from 19 Years of Developing PPs and cPPs (My Take)

- You really have to love this type of work (or be a little crazy) to do it well, because it is very time consuming, very exhausting and very frustrating work
- It is also a long-term time commitment that one has to be willing to make
- Patient is a definite virtue in working on a Technical Community developing PPs/cPPs because nothing happens as quickly as you want it to happen or as smoothly as you want it to happen
- Common Criteria (CC) is very complex, so you need to focus on those parts of the CC that support what you are trying to accomplish – I am still learning things about CC even after 19 years of working with it
- Biggest lesson I learned on the HCD iTC – establish and agree on your procedures and then follow them even when it hurts; every time you don’t you get yourself into self-inflicted trouble

AI Cybersecurity in the EU and US

EU Artificial Intelligence (AI) Act

EU Artificial Intelligence Act Scope

- Providers placing on the market or putting into service AI systems in the Union, irrespective of whether those providers are established within the Union or in a third country
- Users of AI systems located within the Union
- Providers and users of AI systems that are located in a third country, where the output produced by the system is used in the Union
- Not apply to AI systems developed or used exclusively for military purposes
- Not apply to public authorities in a third country nor to international organizations falling within the scope of this Regulation pursuant to paragraph 1, where those authorities or organizations use AI systems in the framework of international agreements for law enforcement and judicial cooperation with the Union or with one or more Member States

EU Artificial Intelligence Act Prohibited Practices

- The placing on the market, putting into service or use of an AI system that deploys subliminal techniques beyond a person’s consciousness in order to materially distort a person’s behavior in a manner that causes or is likely to cause that person or another person physical or psychological harm
- The placing on the market, putting into service or use of an AI system that exploits any of the vulnerabilities of a specific group of persons due to their age, physical or mental disability, in order to materially distort the behaviour of a person pertaining to that group in a manner that causes or is likely to cause that person or another person physical or psychological harm
- The placing on the market, putting into service or use of AI systems by public authorities or on their behalf for the evaluation or classification of the trustworthiness of natural persons over a certain period of time based on their social behavior or known or predicted personal or personality characteristics, with the social score leading to either or both of the following:
  - Detrimental or unfavorable treatment of certain natural persons or whole groups thereof in social contexts which are unrelated to the contexts in which the data was originally generated or collected;
  - Detrimental or unfavorable treatment of certain natural persons or whole groups thereof that is unjustified or disproportionate to their social behaviour or its gravity
-

EU Artificial Intelligence Act
Hi-Risk AI Systems

A Hi-Risk AI System meets both of the following conditions:

- The AI system is intended to be used as a safety component of a product, or is itself a product, covered by the Union harmonization legislation listed in Annex I
- The product whose safety component is the AI system, or the AI system itself as a product, is required to undergo a third-party conformity assessment with a view to the placing on the market or putting into service of that product pursuant to the Union harmonization legislation listed in Annex II

AI Systems can be added to the list of Hi-Risk AI Systems if they meet both of the following conditions:

- The AI systems are intended to be used in any of the areas listed in points 1 to 8 of Annex III
- The AI systems pose a risk of harm to the health and safety, or a risk of adverse impact on fundamental rights, that is, in respect of its severity and probability of occurrence, equivalent to or greater than the risk of harm or of adverse impact posed by the high-risk AI systems already referred to in Annex III

EU Artificial Intelligence Act
Hi-Risk AI Systems

The following are categories of Hi-Risk AI Systems

- Biometric identification and categorization of natural persons:
  - AI systems intended to be used for the 'real-time' and 'post' remote biometric identification of natural persons
- Management and operation of critical infrastructure
  - AI systems intended to be used as safety components in the management and operation of road traffic and the supply of water, gas, heating and electricity
- Education and vocational training:
  - AI systems intended to be used for the purpose of determining access or assigning natural persons to educational and vocational training institutions
  - AI systems intended to be used for the purpose of assessing students in educational and vocational training institutions and for assessing participants in tests commonly required for admission to educational institutions

1. Employment, workers management and access to self-employment:
   - AI systems intended to be used for recruitment or selection of natural persons, notably for advertising vacancies, screening or filtering applications, evaluating candidates in the course of interviews or tests
   - AI intended to be used for making decisions on promotion and termination of work-related contractual relationships, for task allocation and for monitoring and evaluating performance and behavior of persons in such relationships

EU Artificial Intelligence Act
Hi-Risk AI Systems

The following are categories of Hi-Risk AI Systems

- Access to and enjoyment of essential private services and public services and benefits:
  - AI systems intended to be used by public authorities or on behalf of public authorities to evaluate the eligibility of natural persons for public assistance benefits and services, as well as to grant, reduce, revoke, or reclaim such benefits and services
  - AI systems intended to be used to evaluate the creditworthiness of natural persons or establish their credit score, with the exception of AI systems put into service by small scale providers for their own use
  - AI systems intended to be used to dispatch, or to establish priority in the dispatching of emergency first response services, including by firefighters and medical aid
- Law enforcement:

  (a) AI systems intended to be used by law enforcement authorities for making individual risk assessments of natural persons in order to assess the risk of a natural person for offending or reoffending or the risk for potential victims of criminal offences

  (b) AI systems intended to be used by law enforcement authorities as polygraphs and similar tools or to detect the emotional state of a natural person

  (c) AI systems intended to be used by law enforcement authorities to detect deep fakes as referred to in this regulation

  (d) AI systems intended to be used by law enforcement authorities for evaluation of the reliability of evidence in the course of investigation or prosecution of criminal offences

EU Artificial Intelligence Act
Hi-Risk AI Systems Requirements

High-risk AI systems which make use of techniques involving the training of models with data shall be developed on the basis of training, validation and testing data sets that meet the quality criteria:

- Training, validation and testing data sets shall be subject to appropriate data governance and management practices
- Training, validation and testing data sets shall be relevant, representative, free of errors and complete
- Training, validation and testing data sets shall take into account, to the extent required by the intended purpose, the characteristics or elements that are particular to the specific geographical, behavioral or functional setting within which the high risk AI system is intended to be used
- Appropriate data governance and management practices shall apply for the development of high-risk AI systems

EU Artificial Intelligence Act
Other Hi-Risk AI Systems Requirements

Technical documentation of a high-risk AI system shall

- Be drawn up before that system is placed on the market or put into service and shall be kept up-to date
- Provide national competent authorities and notified bodies with all the necessary information to assess the compliance of the AI system with its requirements

Record Keeping shall:

- Be designed and developed with capabilities enabling the automatic recording of events (‘logs’) while the high-risk AI systems is operating
- Logging capabilities shall ensure a level of traceability of the AI system’s functioning throughout its lifecycle that is appropriate to the intended purpose of the system
- Logging capabilities shall provide, at a minimum:
  - Recording of the period of each use of the system (start date and time and end date and time of each use)
  - The reference database against which input data has been checked by the system
  - Input data for which the search has led to a match
  - Identification of the natural persons involved in the verification of the results

EU Artificial Intelligence Act
Other Hi-Risk AI Systems Requirements

- High-risk AI systems shall be designed and developed in such a way to ensure that their operation is sufficiently transparent to enable users to interpret the system’s output and use it appropriately
- High-risk AI systems shall be accompanied by instructions for use in an appropriate digital format or otherwise that include concise, complete, correct and clear information that is relevant, accessible and comprehensible to users

1. High-risk AI systems shall be designed and developed in such a way, including with appropriate human-machine interface tools, that they can be effectively overseen by natural persons during the period in which the AI system is in use
2. Human oversight shall aim at preventing or minimizing the risks to health, safety or fundamental rights that may emerge when a high-risk AI system is used in accordance with its intended purpose or under conditions of reasonably foreseeable misuse
3. Human oversight shall be ensured through either one or all of the following measures:
    - Identified and built, when technically feasible, into the high-risk AI system by the provider before it is placed on the market or put into service
    - Identified by the provider before placing the high-risk AI system on the market or putting it into service and that are appropriate to be implemented by the user

EU Artificial Intelligence Act
Other Hi-Risk AI Systems Requirements

- High-risk AI systems shall be designed and developed in such a way that they achieve, in the light of their intended purpose, an appropriate level of accuracy, robustness and cybersecurity, and perform consistently in those respects throughout their lifecycle
- The levels of accuracy and the relevant accuracy metrics of high-risk AI systems shall be declared in the accompanying instructions of use
- High-risk AI systems shall be resilient as regards errors, faults or inconsistencies that may occur within the system or the environment in which the system operates, in particular due to their interaction with natural persons or other systems
- High-risk AI systems shall be resilient as regards attempts by unauthorized third parties to alter their use or performance by exploiting the system vulnerabilities

EU Artificial Intelligence Act
Requirements on Hi-Risk AI Systems Providers

Providers of high-risk AI systems shall:

- Ensure that their high-risk AI systems are compliant with Hi-Risk AI Systems requirements
- Have a quality management system in place which complies with the AI Act
- Provide the technical documentation of the high-risk AI system
- When under their control, keep the logs automatically generated by their high-risk AI systems
- Ensure that the high-risk AI system undergoes the relevant conformity assessment procedure, prior to its placing on the market or putting into service
- Comply with the registration obligations
- Take the necessary corrective actions, if the high-risk AI system is not in conformity with requirements
- Inform the appropriate national competent authorities of any corrective actions taken
- Affix the marking to their high-risk AI systems to indicate the conformity with the AI Act
- Upon request of a national competent authority, demonstrate the conformity of the high-risk AI system with requirements

EU Artificial Intelligence Act
Requirements on Hi-Risk AI Systems Providers

Providers of high-risk AI systems shall:

- Put a quality management system in place that ensures compliance with this Regulation
- That system shall be documented in a systematic and orderly manner in the form of written policies, procedures and instructions
- Implementation of written policies, procedures and instructions shall be proportionate to the size of the provider's organization
- Draw up the technical documentation that meets the requirements of the AI Act
- Ensure that their systems undergo the relevant conformity assessment procedure in accordance with the AI Act prior to their placing on the market or putting into service
- Keep the logs automatically generated by their high-risk AI systems, to the extent such logs are under their control by virtue of a contractual arrangement with the user or otherwise by law
- Logs shall be kept for a period that is appropriate in the light of the intended purpose of high-risk AI system and applicable legal obligations under Union or national law
- Immediately take the necessary corrective actions if non-conformities are found to bring that system into conformity, to withdraw it or to recall it, as appropriate
- Inform the distributors of the high-risk AI system in question and, where applicable, the authorised representative and importers accordingly

EU Artificial Intelligence Act Requirements on Hi-Risk AI Systems Users

Users of high-risk AI systems shall:

- Use such systems in accordance with the instructions of use accompanying the systems
- To the extent the user exercises control over the input data, ensure that input data is relevant in view of the intended purpose of the high-risk AI system
- Monitor the operation of the high-risk AI system on the basis of the instructions of use
- When they have reasons to consider that the use in accordance with the instructions of use may result in the AI system presenting a risk within the meaning of AI Act, inform the provider or distributor and suspend the use of the system
- Inform the provider or distributor when they have identified any serious incident or any malfunctioning within the meaning of the AI Act and interrupt the use of the AI system
- Keep the logs automatically generated by that high-risk AI system, to the extent such logs are under their control
- Use the information provided under the AI Act to comply with their obligation to carry out a data protection impact assessment

US AI-Related Legislation

AI In Government Act of 2020

AI In Government Act of 2020

Division U of the “Consolidated Appropriations Act, 2021”

Creates AI Center of Excellence (AI CoE) to:

- Facilitate the adoption of artificial intelligence technologies in the Federal Government;
- Improve cohesion and competency in the adoption and use of artificial intelligence within the Federal Government

AI In Government Act of 2020
Duties of the AI Center of Excellence

- Regularly convening individuals from agencies, industry, Federal laboratories, nonprofit organizations, institutions of higher education, and other entities to discuss recent developments in artificial intelligence;
- Collecting, aggregating, and publishing on a publicly available website information regarding programs, pilots, and other initiatives led by other agencies and any other information determined appropriate by the Administrator;
- Advising the Administrator, the Director, and agencies on the acquisition and use of artificial intelligence through technical insight and expertise, as needed;
- Assist agencies in applying Federal policies regarding the management and use of data in applications of artificial intelligence;
- Consulting with agencies, including the Department of Defense, the Department of Commerce, the Department of Energy, the Department of Homeland Security, the Office of Management and Budget, the Office of the Director of National Intelligence, and the National Science Foundation, that operate programs, create standards and guidelines, or otherwise fund internal projects or coordinate between the public and private sectors relating to artificial intelligence;
- Advising the Director on developing policy related to the use of artificial intelligence by agencies
- Advising the Director of the Office of Science and Technology Policy on developing policy related to research and national investment in artificial intelligence

AI In Government Act of 2020
Guidance For Agency Use in AI

No later than 270 days after enactment of this act the Director of the Office of Management and Budget shall issue a memorandum to the head of each agency that shall—

- Inform the development of policies regarding Federal acquisition and use by agencies regarding technologies that are empowered or enabled by artificial intelligence;
- Recommend approaches to remove barriers for use by agencies of artificial intelligence technologies;
- Identify best practices for identifying, assessing, and mitigating any discriminatory impact or bias on the basis of any classification protected under Federal nondiscrimination laws, or any unintended consequence of the use of artificial intelligence; and
- Provide a template of the required contents of the agency plans

AI In Government Act of 2020 Plans

Not later than 180 days after the date on which the Director of the OMB issues the memorandum required under subsection (a) or an update to the memorandum required under subsection (d), the head of each agency shall submit to the Director and post on a publicly available page on the website of the agency:

- (1) a plan to achieve consistency with the memorandum; or
- (2) a written determination that the agency does not use and does not anticipate using artificial intelligence.
- UPDATES.—Not later than 2 years after the date on which the Director of the OMB issues the memorandum required under subsection (a), and every 2 years thereafter for 10 years, the Director of the OMB shall issue updates to the memorandum

National Artificial Intelligence Initiative Act of 2020

National Artificial Intelligence Initiative Act of 2020

Division E, Section 5001 of the “WILLIAM M. (MAC) THORNBERRY NATIONAL DEFENSE AUTHORIZATION ACT FOR FISCAL YEAR 2021”

‘‘artificial intelligence’’ means a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations or decisions influencing real or virtual environments. Artificial intelligence systems use machine and human-based inputs to—(A) perceive real and virtual environments; (B) abstract such perceptions into models through analysis in an automated manner; and (C) use model inference to formulate options for information or action.

National Artificial Intelligence Initiative Act of 2020
National Artificial Intelligence Initiative

Purposes

- (1) Ensure continued United States leadership in artificial intelligence research and development;
- (2) Lead the world in the development and use of trustworthy artificial intelligence systems in the public and private sectors;
- (3) Prepare the present and future United States workforce for the integration of artificial intelligence systems across all sectors of the economy and society; and
- (4) Coordinate ongoing artificial intelligence research, development, and demonstration activities among the civilian agencies, the Department of Defense and the Intelligence Community to ensure that each informs the work of the others

National Artificial Intelligence Initiative Act of 2020
National Artificial Intelligence Initiative

Activities

- (1) Sustain and support for artificial intelligence research and development through grants, cooperative agreements, testbeds, and access to data and computing resources
- (2) Support for K-12 education and postsecondary educational programs, including workforce training and career and technical education programs, and informal education programs
- (3) Support for interdisciplinary research, education, and workforce training programs for students and researchers that promote learning in the methods and systems used in artificial intelligence and foster interdisciplinary perspectives and collaborations among subject matter experts in relevant fields
- (4) Interagency planning and coordination of Federal artificial intelligence research, development, demonstration, standards engagement, and other activities under the Initiative, as appropriate
- (5) Outreach to diverse stakeholders, including citizen groups, industry, and civil rights and disability rights organizations
- (6) Leveraging existing Federal investments to advance objectives of the Initiative
- (7) Support for a network of interdisciplinary artificial intelligence research institutes
- (8) Support opportunities for international cooperation with strategic allies, as appropriate, on the research and development, assessment, and resources for trustworthy artificial intelligence systems

National Artificial Intelligence Initiative Act of 2020 National Artificial Intelligence Research Institutes

National Institute of Standards and Technology (NIST)

- *(1) support measurement research and development of best practices and voluntary standards for trustworthy artificial intelligence systems*
- *(2) produce curated, standardized, representative, high-value, secure, aggregate, and privacy protected data sets for artificial intelligence research, development, and use;*
- *(3) support one or more institutes for the purpose of advancing measurement science, voluntary consensus standards, and guidelines for trustworthy artificial intelligence systems;*
- *(4) support and strategically engage in the development of voluntary consensus standards, including international standards, through open, transparent, and consensus-based processes;*
- *(5) enter into and perform such contracts as may be necessary in the conduct of the work of NIST and on such terms as the Director considers appropriate*
- ***RISK MANAGEMENT FRAMEWORK.—Not later than 2 years after the date of the enactment of this Act, the Director shall work to develop, and periodically update, in collaboration with other public and private sector organizations, including the National Science Foundation and the Department of Energy, a voluntary risk management framework for trustworthy artificial intelligence systems***

National Artificial Intelligence Initiative Act of 2020 National Artificial Intelligence Research Institutes

NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION ARTIFICIAL INTLLIGENCE CENTER

- (1) coordinate and facilitate artificial intelligence research and innovation, tools, systems, and capabilities across the National Oceanic and Atmospheric Administration;
- (2) establish data standards and develop and maintain a central repository for agency-wide artificial intelligence applications;
- (3) accelerate the transition of artificial intelligence research to applications in support of the mission of the National Oceanic and Atmospheric Administration;
- (4) develop and conduct training for the workforce of the National Oceanic and Atmospheric Administration related to artificial intelligence research and application of artificial intelligence for such agency;
- (5) facilitate partnerships between the National Oceanic and Atmospheric Administration and other public sector organizations, private sector organizations, and institutions of higher education for research, personnel exchange, and workforce development with respect to artificial intelligence systems; and
- (6) make data of the National Oceanic and Atmospheric Administration accessible, available, and ready for artificial intelligence applications

National Artificial Intelligence Initiative Act of 2020
National Artificial Intelligence Research Institutes

National Science Foundation (NSF)

- (1) support research, including interdisciplinary research, on artificial intelligence systems and related areas;
- (2) use the existing programs of the National Science Foundation, in collaboration with other Federal departments and agencies, as appropriate to—
  - (A) improve the teaching and learning of topics related to artificial intelligence systems in K-12 education and postsecondary educational programs; and
  - (B) increase participation in artificial intelligence related fields;
- (3) support partnerships among institutions of higher education, Federal laboratories, nonprofit organizations, State, local, and Tribal governments, industry, and potential users of artificial intelligence systems that facilitate collaborative research, personnel exchanges, and workforce development;
- (4) ensure adequate access to research and education infrastructure with respect to artificial intelligence systems;
- (5) conduct prize competitions, as appropriate;
- (6) coordinate research efforts funded through existing programs across the directorates of the National Science Foundation;
- (7) provide guidance on data sharing by grantees to public and private sector organizations consistent with the standards and guidelines developed by NIST; and
- (8) evaluate opportunities for international collaboration with strategic allies on artificial intelligence research and development by

Executive Order 13859
Maintaining American Leadership in Artificial Intelligence

Executive Order 13859 Maintaining American Leadership in Artificial Intelligence

Issued February 11, 2019

Definitions

- ‘artificial intelligence’’ means the full extent of Federal investments in AI, to include: R&D of core AI techniques and technologies; AI prototype systems; application and adaptation of AI techniques; architectural and systems support for AI; and cyberinfrastructure, data sets, and standards for AI; and
- (b) the term ‘‘open data’’ shall, in accordance with OMB Circular A– 130 and memorandum M–13–13, mean ‘‘publicly available data structured in a way that enables the data to be fully discoverable and usable by end users.

Executive Order 13859 Maintaining American Leadership in Artificial Intelligence

Policies and Principles

- Drive development of appropriate technical standards and reduce barriers to the safe testing and deployment of AI technologies
- Train current and future generations of American workers with the skills to develop and apply AI technologies
- Foster public trust and confidence in AI technologies and protect civil liberties, privacy, and American values in their application
- Promote an international environment that supports American AI research and innovation and opens markets for American AI industries, while protecting our technological advantage in AI and protecting our critical AI technologies from acquisition by strategic competitors and adversarial nations

Executive Order 13859 Maintaining American Leadership in Artificial Intelligence - Objectives

- Promote sustained investment in AI R&D in collaboration with industry, academia, international partners and allies, and other non-Federal entities
- Enhance access to high-quality and fully traceable Federal data, models, and computing resources to increase the value of such resources for AI R&D, while maintaining safety, security, privacy, and confidentiality protections consistent with applicable laws and policies
- Reduce barriers to the use of AI technologies to promote their innovative application while protecting American technology, economic and national security, civil liberties, privacy, and values
- Ensure that technical standards minimize vulnerability to attacks from malicious actors and reflect Federal priorities for innovation, public trust, and public confidence in systems that use AI technologies; and develop international standards to promote and protect those priorities
- Train the next generation of American AI researchers and users through apprenticeships; skills programs; and education in science, technology, engineering, and mathematics (STEM), with an emphasis on computer science, to ensure that American workers, including Federal workers, are capable of taking full advantage of the opportunities of AI
- Develop and implement an action plan to protect the advantage of the United States in AI and technology critical to United States economic and national security interests against strategic competitors and foreign adversaries

Executive Order 13859 Maintaining American Leadership in Artificial Intelligence

- Initiative shall be coordinated through the National Science and Technology Council (NSTC) Select Committee on Artificial Intelligence (Select Committee)
- Actions shall be implemented by agencies that conduct foundational AI R&D, develop and deploy applications of AI technologies, provide educational grants, and regulate and provide guidance for applications of AI technologies
- Heads of implementing agencies that also perform or fund R&D (AI R&D agencies), shall consider AI as an agency R&D priority, as appropriate to their respective agencies' missions
- Heads of AI R&D agencies shall budget an amount for AI R&D that is appropriate for this prioritization

Executive Order 13859 Maintaining American Leadership in Artificial Intelligence

- Heads of all agencies shall review their Federal data and models to identify opportunities to increase access and use by the greater non-Federal AI research community in a manner that benefits that community, while protecting safety, security, privacy, and confidentiality
  - In identifying data and models for consideration for increased public access, agencies shall consider issues such as:
    - Privacy and civil liberty protections for individuals who may be affected by increased access and use, as well as confidentiality protections for individuals and other data providers;
    - Safety and security concerns, including those related to the association or compilation of data and models;
    - Data documentation and formatting, including the need for interoperable and machine-readable data formats;
    - Changes necessary to ensure appropriate data and system governance; and
    - Any other relevant considerations

Executive Order 13859
Maintaining American Leadership in Artificial Intelligence

- Heads of implementing agencies that also provide educational grants shall, to the extent consistent with applicable law, consider AI as a priority area within existing Federal fellowship and service programs
- Eligible programs for prioritization shall give preference to American citizens, to the extent permitted by law, and shall include:
    - High school, undergraduate, and graduate fellowship; alternative education; and training programs;
    - Programs to recognize and fund early-career university faculty who conduct AI R&D, including through Presidential awards and recognitions;
    - Scholarship for service programs;
    - Direct commissioning programs of the United States Armed Forces; and
    - Programs that support the development of instructional programs and curricula that encourage the integration of AI technologies into courses in order to facilitate personalized and adaptive learning experiences for formal and informal education and training

Executive Order 13960
Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government

Executive Order 13960
Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government

Issued December 3, 2020

Policy

- Promote the innovation and use of AI, where appropriate, to improve Government operations and services in a manner that fosters public trust, builds confidence in AI, protects our Nation’s values, and remains consistent with all applicable laws, including those related to privacy, civil rights, and civil liberties
- Responsible agencies shall, when considering the design, development, acquisition, and use of AI in Government, be guided by the common set of Principles which are designed to foster public trust and confidence in the use of AI, protect our Nation’s values, and ensure that the use of AI remains consistent with all applicable laws, including those related to privacy, civil rights, and civil liberties

Executive Order 13960
Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government

Policy

- When designing, developing, acquiring, and using AI in the Federal Government, agencies shall adhere to the following Principles:
  - Lawful and respectful of our Nation’s values (including those addressing privacy, civil rights, and civil liberties)
  - Purposeful and performance-driven
  - Accurate, reliable, and effective
  - Safe, secure, and resilient
  - Understandable
  - Responsible and traceable
  - Regularly monitored
  - Transparent
  - Accountable

Executive Order 13960 Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government

- The Principles and implementation guidance in this order shall apply to AI designed, developed, acquired, or used specifically to advance the execution of agencies' missions, enhance decision making, or provide the public with a specified benefit
- This order applies to both existing and new uses of AI; both standalone AI and AI embedded within other systems or applications; AI developed both by the agency or by third parties on behalf of agencies for the fulfilment of specific agency missions, including relevant data inputs used to train AI and outputs used in support of decision making; and agencies' procurement of AI applications
- This order does not apply to:
  - AI used in defense or national security systems, in whole or in part, although agencies shall adhere to other applicable guidelines and principles for defense and national security purposes, such as those adopted by the Department of Defense and the Office of the Director of National Intelligence;
  - AI embedded within common commercial products, such as word processors or map navigation systems, while noting that Government use of such products must nevertheless comply with applicable law and policy to assure the protection of safety, security, privacy, civil rights, civil liberties, and American values; and
  - AI research and development (R&D) activities, although the Principles and OMB implementation guidance should inform any R&D directed at potential future applications of AI in the Federal Government

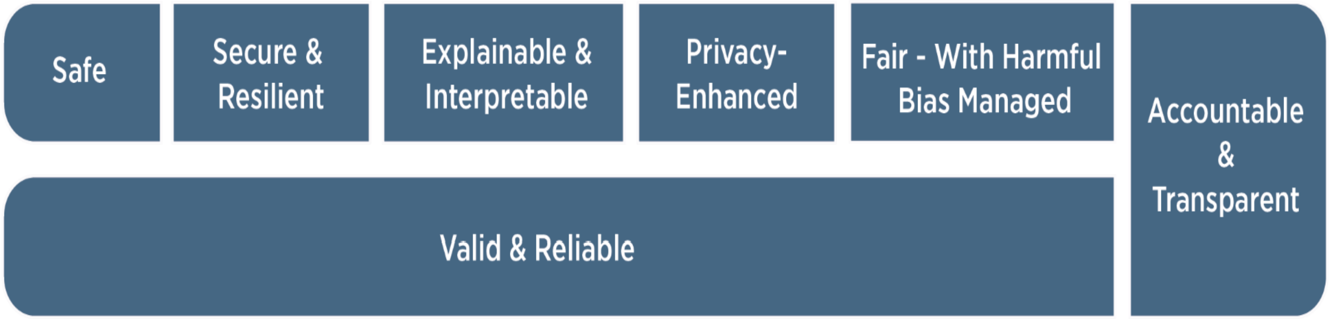
NIST AI 100.1
NIST AI Risk Management Framework
(AI RMF 1.0)

NIST AI 100-1
NIST AI Risk Management Framework

- Published January 2023
- Offers a resource to the organizations designing, developing, deploying, or using AI systems to help manage the many risks of AI and promote trustworthy and responsible development and use of AI systems.
- Is intended to be:
  - ***Voluntary***, rights-preserving, non-sector-specific, and use-case agnostic, providing flexibility to organizations of all sizes and in all sectors and throughout society to implement the approaches in the Framework
  - Practical, to adapt to the AI landscape as AI technologies continue to develop, and to be operationalized by organizations in varying degrees and capacities so society can benefit from AI while also being protected from its potential harms
  - Flexible and to augment existing risk practices which should align with applicable laws, regulations, and norms
- Is designed to equip organizations and individuals – referred to here as *AI actors* – with approaches that increase the trustworthiness of AI systems, and to help foster the responsible design, development, deployment, and use of AI systems over time
- Offers approaches to minimize anticipated negative impacts of AI systems *and* identify opportunities to maximize positive impacts
- Designed to address new risks as they emerge

NIST AI 100-1
NIST AI Risk Management Framework
Characteristics of Trustworthiness

Trustworthy AI is: valid and reliable, safe, secure and resilient, accountable and transparent, explainable and interpretable, privacy-enhanced, and *Fair with Harmful Bias Managed*



Safe

Secure & Resilient

Explainable & Interpretable

Privacy-Enhanced

Fair - With Harmful Bias Managed

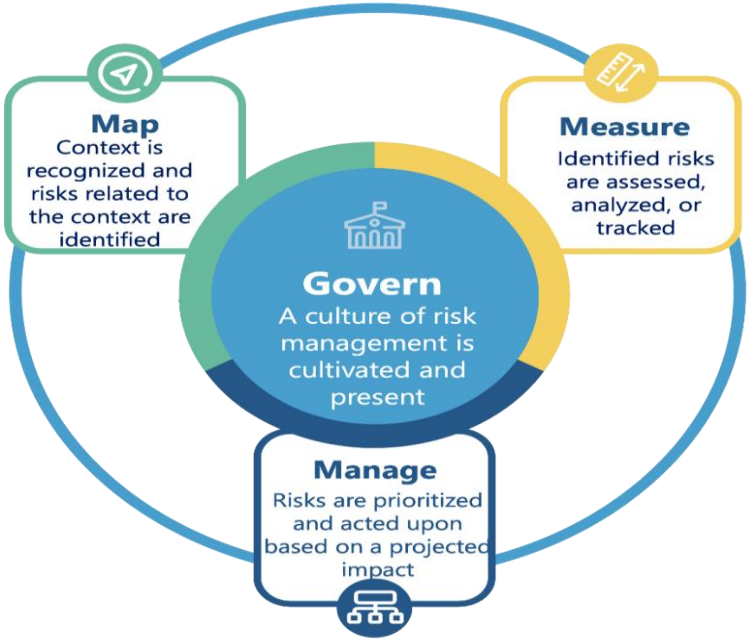
Accountable & Transparent

Valid & Reliable

- Approaches which enhance AI trustworthiness can also contribute to a reduction of AI risks
- Addressing AI trustworthy characteristics individually will not assure AI system trustworthiness, and tradeoffs are always involved
- Increasing the breadth and diversity of stakeholder input throughout the AI lifecycle can enhance opportunities for identifying AI system benefits and positive impacts, and increase the likelihood that risks arising in social contexts are managed appropriately

NIST AI 100-1
NIST AI Risk Management Framework
Core Functions

- **Govern**: Cultivate and implement a culture of risk management within organizations developing, deploying, or acquiring AI systems
- **Map**: Establish the context to frame risks related to an AI system
- **Measure**: Employ quantitative, qualitative, or mixed-method tools, techniques, and methodologies to analyze, assess, benchmark, and monitor AI risk and related impacts
- **Manage**: Entails allocating risk management resources to mapped and measured risks on a regular basis and as defined by the Govern function


Map
Context is recognized and risks related to the context are identified
Measure
Identified risks are assessed, analyzed, or tracked
Govern
A culture of risk management is cultivated and present
Manage
Risks are prioritized and acted upon based on a projected impact

NIST AI 100-1
NIST AI Risk Management Framework
AI-Specific Risks

- The data used for building an AI system may not be a true or appropriate representation of the context or intended use of the AI system, and the ground truth may either not exist or not be available
- Harmful bias and other data quality issues can affect AI system trustworthiness, which could lead to negative impacts
- AI system dependency and reliance on data for training tasks, combined with increased volume and complexity typically associated with such data
- Intentional or unintentional changes during training may fundamentally alter AI system performance
- Datasets used to train AI systems may become detached from their original and intended context or may become stale or outdated relative to deployment context
- AI system scale and complexity (many systems contain billions or even trillions of decision points) housed within more traditional software applications
- Use of pre-trained models that can advance research and improve performance can also increase levels of statistical uncertainty and cause issues with bias management, scientific validity, and reproducibility.
- Higher degree of difficulty in predicting failure modes for emergent properties of large-scale pre-trained models

NIST AI 100-1
NIST AI Risk Management Framework
AI-Specific Risks

- Privacy risk due to enhanced data aggregation capability for AI systems
- AI systems may require more frequent maintenance and triggers for conducting corrective maintenance due to data, model, or concept drift
- Increased opacity and concerns about reproducibility
- Underdeveloped software testing standards and inability to document AI-based practices to the standard expected of traditionally engineered software for all but the simplest of cases.
- Difficulty in performing regular AI-based software testing, or determining what to test, since AI systems are not subject to the same controls as traditional code development
- Computational costs for developing AI systems and their impact on the environment and planet
- Inability to predict or detect the side effects of AI-based systems beyond statistical measures
- **Social and ethical impact of the use of AI systems**

HCD Security Guidelines

Liaison Status

Trusted Computing Group (TCG)

- **Recent and Next TCG Members Meetings**
  - **TCG Hybrid F2F (Berlin, Germany) – 27-29 June 2023 – Ira called in**
  - **TCG Hybrid F2F (Kirkland, WA) – 24-26 October 2023 – Ira to call in**
- **Trusted Mobility Solutions (TMS) – Ira is co-chair and co-editor**
  - **Formal Liaisons – GP (TEE, SE, TPS), ETSI (NFV/MEC/SAI Security and Privacy)**
  - **Informal Liaisons – 3GPP, GSMA, IETF, ISO, ITU-T, SAE, US NIST**
  - ***TCG TMS Use Cases v2 – published September 2018***
- **Mobile Platform (MPWG) – Ira is co-editor**
  - **Formal and Informal Liaisons – jointly with TMS WG above**
  - ***TCG Mobile Reference Architecture v2 – publication approved by TC July 2023***
  - ***TCG MARS 1.0 Mobile Profile – new work-in-progress Q4 2021***
  - ***TCG TPM 2.0 Mobile Common Profile – work-in-progress deferred to Q4 2023***
  - ***TCG Runtime Integrity Preservation for Mobile Devices – published Nov 2019***
  - ***GP TPS Client API / Entity Attestation Protocol / COSE Keystore – joint work***
- **Recent Specifications**
  - **http://www.trustedcomputinggroup.org/resources**
  - ***TCG Mobile Reference Architecture v2 – publication approved by TC July 2023***
  - ***TCG PC Client Platform Firmware Profile v1.06 – public review July 2023***
  - ***TCG Algorithm Registry v1.34 – public review June 2023***
  - ***TCG Component Class Registry v1r14 – published May 2023***
  - ***TCG MARS API v1 – published May 2023***
  - ***TCG Measurement and Attestation RootS (MARS) Library – published January 2023***

Internet Engineering Task Force (IETF) (1 of 4)

- **Recent and Next IETF Members Meetings**
  - **IETF 117 Hybrid F2F (San Francisco, CA) – 24-28 July 2023 – Ira called in**
  - **IETF 118 Hybrid F2F (Prague, Czech Republic) – 6-10 November 2023 – Ira to call in**
  - **IETF 119 Hybrid F2F (Brisbane, Australia) – 18-22 March 2024 – Ira to call in**
- **Transport Layer Security (TLS)**
  - **IETF Delegated Credentials for TLS and DTLS – RFC 9345 – July 2023**
    https://datatracker.ietf.org/doc/rfc9345/
  - **IETF Exported Authenticators in TLS – RFC 9261 – July 2022**
    https://datatracker.ietf.org/doc/rfc9261/
  - **IETF Importing External Pre-Shared Keys (PSKs) for TLS 1.3 – RFC 9258 – July 2022**
    https://datatracker.ietf.org/doc/rfc9258/
  - **IETF Guidance for External Pre-Shared Key (PSK) Usage in TLS – RFC 9257 – July 2022**
    https://datatracker.ietf.org/doc/rfc9257/
  - **IETF DTLS Protocol Version 1.3 – RFC 9147 – April 2022**
    https://datatracker.ietf.org/doc/rfc9147/
  - **IETF SSLKEYLOGFILE Format for TLS – draft-01 – July 2023**
    https://datatracker.ietf.org/doc/draft-thomson-tls-keylogfile/
  - **IETF Flags Extension for TLS 1.3 – draft-12 – July 2023**
    https://datatracker.ietf.org/doc/draft-ietf-tls-tlsflags/
  - **IETF TLS 1.3 – draft-09 – July 2023**
    https://datatracker.ietf.org/doc/draft-ietf-tls-rfc8446bis/
  - **IETF Well-known URL for publishing ECHConfigList values – draft-03 – July 2023**
    https://datatracker.ietf.org/doc/draft-ietf-tls-wkech/
  - **IETF Abridged Compression for WebPKI Certificates – draft-00 – July 2023**
    https://datatracker.ietf.org/doc/draft-jackson-tls-cert-abridge/
  - **IETF TLS 1.2 is Frozen – draft-01 – June 2023**
    https://datatracker.ietf.org/doc/draft-rsalz-tls-tls12-frozen/
  - **IETF Post-quantum hybrid ECDHE-Kyber Key Agreement for TLSv1.3 – draft-01 – May 2023**
    https://datatracker.ietf.org/doc/draft-kwiatkowski-tls-ecdhe-kyber/

Internet Engineering Task Force (IETF) (2 of 4)

- **Security Automation & Continuous Monitoring (SACM) – Closed**
  - **IETF Concise Software Identifiers – RFC 9393 – June 2023**
    https://datatracker.ietf.org/doc/rfc9393/
  - **IETF Software Inventory Message and Attributes for PA-TNC – RFC 8412 – July 2018**
    https://datatracker.ietf.org/doc/rfc8412/
- **Concise Binary Object Representation (CBOR)**
  - **IETF Stable Storage for Items in CBOR – RFC 9277 – August 2022**
    https://datatracker.ietf.org/doc/rfc9277/
  - **IETF Additional Control Ops for CDDL – RFC 9165 – December 2021**
    https://datatracker.ietf.org/doc/rfc9165/
  - **IETF CBOR tags for IPv4/v6 Adresses – RFC 9164 – December 2021**
    https://datatracker.ietf.org/doc/rfc9164/
  - **IETF CBOR Tags for OIDs – RFC 9090 – July 2021**
    https://datatracker.ietf.org/doc/rfc9090/
  - **IETF App-Oriented Literals in CBOR Ext Diag Notation – draft-02 – July 2023**
    https://datatracker.ietf.org/doc/draft-ietf-cbor-edn-literals/
  - **IETF CBOR Tags for Time, Duration, and Period – draft-09 – July 2023**
    https://datatracker.ietf.org/doc/draft-ietf-cbor-time-tag/
  - **IETF Packed CBOR – draft-09 – July 2023**
    https://datatracker.ietf.org/doc/draft-ietf-cbor-packed/
  - **IETF CDDL Module Structure – draft-00 – June 2023**
    https://datatracker.ietf.org/doc/draft-ietf-cbor-cddl-modules/
  - **IETF Updates to the CDDL grammar of RFC 8610 – draft-00 – June 2023**
    https://datatracker.ietf.org/doc/draft-ietf-cbor-update-8610-grammar/
  - **IETF More Control Operators for CDDL – draft-00 – June 2023**
    https://datatracker.ietf.org/doc/draft-ietf-cbor-cddl-more-control/

Internet Engineering Task Force (IETF) (3 of 4)

- **Remote ATtestation ProcedureS (RATS)**
  - **IETF RATS Architecture – RFC 9334 – January 2023**
    https://datatracker.ietf.org/doc/rfc9334/
  - **IETF EAT Media Types – draft-04 – July 2023**
    https://datatracker.ietf.org/doc/draft-ietf-rats-eat-media-type/
  - **IETF Concise Reference Integrity Manifest (CoRIM) – draft-02 – July 2023**
    https://datatracker.ietf.org/doc/draft-ietf-rats-corim/
  - **IETF EAT-based Key Attestation Token – draft-01 – July 2023**
    https://datatracker.ietf.org/doc/draft-bft-rats-kat/
  - **IETF Epoch Markers – draft-05 – July 2023**
    https://datatracker.ietf.org/doc/draft-birkholz-rats-epoch-markers/
  - **IETF RATS Endorsements – draft-02 – July 2023**
    https://datatracker.ietf.org/doc/draft-dthaler-rats-endorsements/
  - **IETF EAT Attestation Results – draft-01 – July 2023**
    https://datatracker.ietf.org/doc/draft-fv-rats-ear/
  - **IETF ARM Platform Security Architecture Attestation Token – draft-12 – July 2023**
    https://datatracker.ietf.org/doc/draft-tschofenig-rats-psa-token/
  - **IETF Entity Attestation Token (EAT) – draft-21 – June 2023**
    https://datatracker.ietf.org/doc/draft-ietf-rats-eat/
  - **IETF Intel Profile for CoRIM – draft-00 – June 2023**
    https://datatracker.ietf.org/doc/draft-cds-rats-intel-corim-profile/
  - **IETF RATS Conceptual Messages Wrapper – draft-03 – June 2023**
    https://datatracker.ietf.org/doc/draft-ftbs-rats-msg-wrap/

Internet Engineering Task Force (IETF) (4 of 4)

- **IRTF Crypto Forum Research Group (CFRG) – future algorithms**
  - **IRTF Hybrid Public Key Encryption – RFC 9180 – February 2022**
    https://datatracker.ietf.org/doc/rfc9180/
  - **IRTF Argon2 password hash and proof-of-work – RFC 9106 – September 2021**
    https://datatracker.ietf.org/doc/rfc9106/
  - **IRTF AEGIS family of authenticated encryption algorithms – draft-04 – July 2023**
    https://datatracker.ietf.org/doc/draft-irtf-cfrg-aegis-aead/
  - **IRTF CPace, a balanced composable PAKE – draft-08 – July 2023**
    https://datatracker.ietf.org/doc/draft-irtf-cfrg-cpace/
  - **IRTF Key Blinding for Signature Schemes – draft-04 – July 2023**
    https://datatracker.ietf.org/doc/draft-irtf-cfrg-signature-key-blinding/
  - **IRTF Secp256k1-based DHKEM for HPKE – draft-00 – July 2023**
    https://datatracker.ietf.org/doc/draft-wahby-cfrg-hpke-kem-secp256k1/
  - **IRTF Merkle Tree Ladder Mode (MTL) Signatures – draft-00 – July 2023**
    https://datatracker.ietf.org/doc/draft-harvey-cfrg-mtl-mode/
  - **IRTF BBS Signature Scheme – draft-03 – July 2023**
    https://datatracker.ietf.org/doc/draft-irtf-cfrg-bbs-signatures/
  - **IRTF Guidelines for Writing Cryptography Specifications – draft-00 – July 2023**
    https://datatracker.ietf.org/doc/draft-irtf-cfrg-cryptography-specification/
  - **IRTF Two-Round Threshold Schnorr Sigs with FROST – draft-14 – July 2023**
    https://datatracker.ietf.org/doc/draft-irtf-cfrg-frost/
  - **IRTF RSA Blind Signatures – draft-14 – July 2023**
    https://datatracker.ietf.org/doc/draft-irtf-cfrg-rsa-blind-signatures/
  - **IRTF Deterministic Nonce-less Hybrid Public Key Encryption – draft-01 – July 2023**
    https://datatracker.ietf.org/doc/draft-irtf-cfrg-dnhpke/

Next Steps – IDS WG

- Next IDS WG Meeting– August 24, 2023
- Next IDS Face-to-Face Meeting likely November 16, 2023 at PWG November 2023 F2F
- Start looking at involvement in some of these other standards activities individually and maybe as a WG

Full slide sets for the Special Topic items can be found at the following:

EU AI Act: https://ftp.pwg.org/pub/pwg/ids/Presentation/AI Act.pdf

US AI Legislation: https://ftp.pwg.org/pub/pwg/ids/Presentation/US AI Legislation.pdf

NIST AI Risk Management Framework: https://ftp.pwg.org/pub/pwg/ids/Presentation/NIST AI Risk Management Framework v2.pdf

Backup

HCD iTC
Issues Post-Version 1.0 – CNSA 2.0

- Commercial National Security Algorithm (CNSA) 2.0 released by NSA Sep 2022
- Addresses problem that future deployment of a cryptanalytically relevant quantum computer (CRQC) would break public-key systems still used today
- Need to plan, prepare, and budget for an effective transition to quantum-resistant (QR) algorithms, to assure continued protection of National Security Systems (NSS) and related assets
- Is an update to CNSA 1.0 Algorithms
- Applies to all NSS use of public cryptographic algorithms (as opposed to algorithms NSA developed), including those on all unclassified and classified NSS
- Using any cryptographic algorithms the National Manager did not approve is generally not allowed, and requires a waiver specific to the algorithm, implementation, and use case
- Per CNSSP 11, software and hardware providing cryptographic services require NIAP or NSA validation in addition to meeting the requirements of the appropriate version of CNSA

Commercial National Security Algorithm (CNSA) Suite 2.0 Algorithms

Algorithm	**Function**	**Specification**	**Parameters**
Advanced Encryption Standard (AES)	Symmetric block cipher for information protection	FIPS PUB 197	Use 256-bit keys for all classification levels
CRYSTALS-Kyber	Asymmetric algorithm for key establishment	TBD	Use Level V parameters for all classification levels
CRYSTALS-Dilithium	Asymmetric algorithm for digital signatures	TBD	Use Level V parameters for all classification levels
Secure Hash Algorithm (SHA)	Algorithm for computing a condensed representation of information	FIPS PUB 180-4	Use SHA-384 or SHA-512 for all classification levels
Leighton-Micali Signature (LMS)	Asymmetric algorithm for digitally signing firmware and software	NIST SP 800-208	All parameters approved for all classification levels SHA256/192 recommended
Xtended Merkle Signature Scheme (XMSS)	Asymmetric algorithm for digitally signing firmware and software	NIST SP 800-208	All parameters approved for all classification levels

Transitioning to CNSA Suite 2.0

- The timing of the transition depends on the proliferation of standards-based implementations
- NSA expects the transition to QR algorithms for NSS to be complete by 2035 in line with NSM-10.
- NSA urges vendors and NSS owners and operators to make every effort to meet this deadline.
- Where feasible, NSS owners and operators will be required to prefer CNSA 2.0 algorithms when configuring systems during the transition period.
- When appropriate, use of CNSA 2.0 algorithms will be mandatory in classes of commercial products within NSS, while reserving the option to allow other algorithms in specialized use cases

Detailed NIAP Transition Plan for CNSA Suite 2.0

- Currently all NIAP PPs must have CNSA 1.0 algorithms
- Will add SHA-512 to all NIAP PPs
- Will require either CNSA 1.0 or CNSA 2.0 be mandatory on all NIAP PPs
- Will implement CNSA asymmetric algorithms for software/firmware signing per following
  - LMS – 1H 2023
  - XMSS – 2H 2023
- Will implement following Key Establishment CNSA 2.0 algorithms in all NIAP PPs when they are standardized and all relevant Assurance Activities have been defined and agreed upon:
  - CRYSTALS - Kyber
  - CRYSTALS – Dilithium (used for Digital Signatures)
- Will deprecate CNSA 1.0 in 2030 – 2033 timeframe
- No current timeline established to make CNSA 2.0 mandatory
  - Will make use of CNSA 2.0 mandatory to be listed on PCL at some point
- Will work with vendors to help try to meet NSA schedule
- Will discuss with CCRA and engage with iTCs how best to integrate CNSA 2.0 into cPPs