IDS WG Meeting Minutes
January 26, 2023

This IDS WG Meeting was started at approximately 3:00 pm ET on January 26, 2023.

Attendees

Matt Glockner	Dodson
Graydon Dodson	Lexmark
Jeremy Leber	Lexmark
Ira McDonald	High North
Alan Sukert	
Bill Wagner	TIC
Steve Young	Canon

Agenda Items

1. The topics to be covered during this meeting were:
   - Special topics on the EU Artificial Intelligence Act and the NIAP Artificial Intelligence Risk Management Framework
   - Discussion on what will be covered in the upcoming IDS Face-to-Face Meeting on February 9th
   - Round Table
2. Meeting began by stating the PWG Anti-Trust Policy which can be found at https://www.pwg.org/chair/membership_docs/pwg-antitrust-policy.pdf and the PWG Intellectual Property Policy which can be found at https://www.pwg.org/chair/membership_docs/pwg-ip-policy.pdf.
3. Al presented the first of the meeting's week's special topics on the "REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS**",** or simply known as the EU AI Act which is what it will be referred to for the rest of these minutes. The slides Al used can be found at https://ftp.pwg.org/pub/pwg/ids/Presentation/AI Act.pdf.

   The main items covered in the presentation were:

   - The purposes of the EU AI Act where to establish:
     - Harmonized rules for the placing on the market, the putting into service and the use of artificial intelligence systems ('AI systems') in the Union
     - Prohibit certain artificial intelligence practices
     - Specific requirements for high-risk AI systems and obligations for operators of such systems
     - Harmonize transparency rules for AI systems intended to interact with natural persons, emotion recognition systems and biometric categorization systems, and AI systems used to generate or manipulate image, audio or video content
     - Rules on market monitoring and surveillance

     Al noted that the concept of "harmonization" is very important in the EU. "Harmonization" in this context really refers to ensuring each of the EU regulations is consistent with all other applicable EU regulations in terms of requirements, references, etc.
   - The scope of the EU AI Act is:
     - 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

AI systems developed or used exclusively for military purposes are exempt from this regulation, which is typical of this type of regulation

- Some key definitions that are necessary to understand the AI Act are:
  - **'artificial intelligence system' (AI system)**: Software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with
  - **'provider'**: A natural or legal person, public authority, agency or other body that develops an AI system or that has an AI system developed with a view to placing it on the market or putting it into service under its own name or trademark, whether for payment or free of charge
  - **'user'**: Any natural or legal person, public authority, agency or other body using an AI system under its authority, except where the AI system is used in the course of a personal non-professional activity
  - **'authorized representative'**: Any natural or legal person established in the Union who has received a written mandate from a provider of an AI system to, respectively, perform and carry out on its behalf the obligations and procedures established by this Regulation
  - **'importer'**: Any natural or legal person established in the Union that places on the market or puts into service an AI system that bears the name or trademark of a natural or legal person established outside the Union
  - **'distributor'**: Any natural or legal person in the supply chain, other than the provider or the importer, that makes an AI system available on the Union market without affecting its properties
  - **'notifying authority'**: The national authority responsible for setting up and carrying out the necessary procedures for the assessment, designation and notification of conformity assessment bodies and for their monitoring
  - **'conformity assessment'**: The process of verifying whether the requirements set out in the EU AI Act relating to an AI system have been fulfilled
  - **'conformity assessment body'**: A body that performs third-party conformity assessment activities, including testing, certification and inspection
  - **'notified body'**: A conformity assessment body designated in accordance with this Regulation and other relevant Union harmonization legislation
  - **'national supervisory authority'**: The authority to which a Member State assigns the responsibility for the implementation and application of this Regulation, for coordinating the activities entrusted to that Member State, for acting as the single contact point for the Commission, and for representing the Member State at the European Artificial Intelligence Board
  - **'national competent authority'**: The national supervisory authority, the notifying authority and the market surveillance authority

  Al noted that the EU has a lot of "bodies: that are formed in a hierarchical manner associated with every Regulation.

- The EU AI Act prohibits the following:
  - 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 behavior 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 behavior or its gravity
  - The use of ‘real-time’ remote biometric identification systems in publicly accessible spaces for the purpose of law enforcement, unless and in as far as such use is strictly necessary for one of the following objectives:
    - the targeted search for specific potential victims of crime, including missing children
    - the prevention of a specific, substantial and imminent threat to the life or physical safety of natural persons or of a terrorist attack
    - the detection, localization, identification or prosecution of a perpetrator or suspect of a criminal offence and punishable in the Member State concerned by a custodial sentence or a detention order for a maximum period of at least three years, as determined by the law of that Member State
  - The use of ‘real-time’ remote biometric identification systems in publicly accessible spaces for the purpose of law enforcement for any of the objectives referred to in paragraph 1 point d) shall take into account the following elements:
    - the nature of the situation giving rise to the possible use, in particular the seriousness, probability and scale of the harm caused in the absence of the use of the system
    - the nature of the situation giving rise to the possible use, in particular the seriousness, probability and scale of the harm caused in the absence of the use of the system

  What all these prohibitions have in common is that they all basically revolve around the concept of “do no harm” to the human users of an AI System. The only exception, which is found throughout this regulation, is when ‘real-time’ remote biometric identification systems are used for law enforcement purposed or to prevent terrorist and related attacks on EU members.

- The main focus of the EU AI Act is on Hi-Risk AI Systems, which are defined as AI Systems 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 of the EU AI Act (Note: Annex ! addresses “AI Techniques and Approaches”)
  - 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 of the EU AI Act (Note: Annex !I is a list of EU Regulations that the EU AI Act must harmonize with – it is two full pages of EU Regulations which shows how big the issue of harmonization is within the EU).

  AI systems that could “pose a risk of harm to the health and safety, or a risk of adverse impact on fundamental rights” can be added to the list of Hi-Risk AI Systems.

  The categories and types of Hi-Risk AI Systems are:

  - 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
- 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
- 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:
  - 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
  - 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
  - AI systems intended to be used by law enforcement authorities to detect deep fakes as referred to in this regulation
  - 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
  - AI systems intended to be used by law enforcement authorities for predicting the occurrence or reoccurrence of an actual or potential criminal offence based on profiling of natural persons as referred to in Article 3(4) of Directive (EU) 2016/680 or assessing personality traits and characteristics or past criminal behavior of natural persons or groups
  - AI systems intended to be used by law enforcement authorities for profiling of natural persons as referred to in Article 3(4) of Directive (EU) 2016/680 in the course of detection, investigation or prosecution of criminal offences
  - AI systems intended to be used for crime analytics regarding natural persons, allowing law enforcement authorities to search complex related and unrelated large data sets available in different data sources or in different data formats in order to identify unknown patterns or discover hidden relationships in the data
- Administration of justice and democratic processes:
  - AI systems intended to assist a judicial authority in researching and interpreting facts and the law and in applying the law to a concrete set of facts

What these categories have in common is that they all are in areas that directly impact users or are used for law enforcement purposes.

- What was interesting is that the EU AI Act requires that the requirements for Hi-Risk AI Systems have to be connected with a Risk Management System. The EU AI Act requires that this Risk Management System:
  - Shall be established, implemented, documented and maintained in relation to high-risk AI systems
  - Shall consist of a continuous iterative process run throughout the entire lifecycle of a high-risk AI system, requiring regular systematic updating
  - Shall comprise the following:
    - Identification and analysis of the known and foreseeable risks associated with each high-risk AI system
    - Estimation and evaluation of the risks that may emerge when the high-risk AI system is used in accordance with its intended purpose and under conditions of reasonably foreseeable misuse
    - Evaluation of other possibly arising risks based on the analysis of data gathered from the post-market monitoring system
    - Adoption of suitable risk management measures

  These are typical requirements for a risk management system, as was shown when Al did a special topic on the NIST Risk Management Framework at a previous IDS WG meeting.
- The EU AI Act, like all similar EU Regulations, places a heavy emphasis on requirements for metrics and measures to assess how well the requirements in the regulation are being met. For example, the EU AI Act has the following metrics requirements for the risk management system:
  - Elimination or reduction of risks as far as possible through adequate design and development
  - Where appropriate, implementation of adequate mitigation and control
  - Measures in relation to risks that cannot be eliminated
  - Measures shall enable the individuals to whom human oversight is assigned to do the following, as appropriate to the circumstances:
    - Fully understand the capacities and limitations of the high-risk AI system and be able to duly monitor its operation, so that signs of anomalies, dysfunctions and unexpected performance can be detected and addressed as soon as possible
    - Remain aware of the possible tendency of automatically relying or over-relying on the output produced by a high-risk AI system (‘automation bias’), in particular for high-risk AI systems used to provide information or recommendations for decisions to be taken by natural persons
    - Be able to correctly interpret the high-risk AI system’s output, taking into account in particular the characteristics of the system and the interpretation tools and methods available
    - Be able to decide, in any particular situation, not to use the high-risk AI system or otherwise disregard, override or reverse the output of the high-risk AI system
    - Be able to intervene on the operation of the high-risk AI system or interrupt the system through a “stop” button or a similar procedure

  What is interesting in these metrics requirements, and is a theme in the EU AI Act, is the importance of human oversite of a Hi-Risk AI System.
- The requirements for testing of a Hi-Risk AI System are typical test requirements. Specifically,
  - High-risk AI systems shall be tested for the purposes of identifying the most appropriate risk management measures.

  - Testing shall ensure that high-risk AI systems perform consistently for their intended purpose and they are in compliance with the requirements
  - Testing procedures shall be suitable to achieve the intended purpose of the AI system and do not need to go beyond what is necessary to achieve that purpose
  - The testing of the high-risk AI systems shall be performed, as appropriate, at any point in time throughout the development process, and, in any event, prior to the placing on the market or the putting into service
  - Testing shall be made against preliminarily defined metrics and probabilistic thresholds that are appropriate to the intended purpose of the high-risk AI system

  This last one is interesting in that it explicitly calls out testing of "probabilistic thresholds"; that is unusual in test requirements.

- There are also requirements governing training of Hi-Risk AI Systems (see Slide 15). The two main requirements are that training, validation and testing data sets shall be (1) subject to appropriate data governance and management practices and (2) relevant, representative, free of errors and complete. How this second requirement can or will be assessed for compliance will be a challenge,

- There are also requirements (see Slide 16) for technical documentation of a high-risk AI system to be drawn up before that system is placed on the market or put into service and be kept up-to date and for record-keeping in the form of logs – it is not clear whether this is referring to paper logs, audit logs, or some other type of "log". The record-keeping requirements include enablement of automatic recording of events while the high-risk AI systems is operating.

  Record-keeping is required to collect, as 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

  Ira noted that the log requirements in the EU AI Act and requirements around incidents of misbehavior are useless because an AI system is basically a "black box", so you really can't get any real information on what is going on inside the "black box".

- Slides 17 and 18 list other requirements for high-risk AI systems that are related to human oversight and design and development of high-risk AI systems. A couple of key ones are:

  - 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
  - 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
  - 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

  Resiliency is another major theme in the EU AI Act – how resiliency will be evaluated will be another challenge.

- Slides 19-21 provide a long list of requirements for the providers of high-risk AI systems. Many of these requirements are what would be expected - like “ensure that their high-risk AI systems are compliant with Hi-Risk AI Systems requirements or prior to its placing on the market or putting into service). A few of the more interesting beyond these expected ones are:
  - Have a quality management system in place which complies with the AI Act
  - When under their control, keep the logs automatically generated by their high-risk AI systems
  - Comply with the registration obligations
  - 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
  - That system shall be documented in a systematic and orderly manner in the form of written policies, procedures and instructions and shall be proportionate to the size of the provider’s organization
  - 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
  - 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
  - Inform the distributors of the high-risk AI system in question and, where applicable, the authorized representative and importers accordingly
  - Where the high-risk AI system presents a risk and that risk is known to the provider of the system, immediately inform the national competent authorities of the Member States in which it made the system available and, where applicable, the notified body that issued a certificate for the high-risk AI system of the non-compliance and of any corrective actions taken
- When Al gave the presentation he realized he was missing a slide of the requirements for users of high-risk AI systems. These requirements are:
  - 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
- Transparency is another major theme of this regulation, where “transparency” is related to providing proper and necessary visibility into the execution of this regulation. The specific requirements are:
  - Providers shall ensure that AI systems intended to interact with natural persons are designed and developed in such a way that natural persons are informed that they are interacting with an AI system, unless this is obvious from the circumstances and the context of use
  - Users of an emotion recognition system or a biometric categorization system shall inform of the operation of the system the natural persons exposed thereto
  - Does not apply to AI systems used for biometric categorization which are permitted by law to detect, prevent and investigate criminal offences

    - Users of an AI system that generates or manipulates image, audio or video content that appreciably resembles existing persons, objects, places or other entities or events and would falsely appear to a person to be authentic or truthful ('deep fake'), shall disclose that the content has been artificially generated or manipulated

  Al noted the last requirement dealing with false images and "deep fakes" – this is a big issue in the US as well, so it was nice to see those requirements in this regulation. However, this particular requirement does not apply where the use is authorized by law to detect, prevent, investigate and prosecute criminal offences or for the exercise of the right to freedom of expression and the right to freedom of the arts and sciences.

- Slide 24 gives a list of some of the other chapters in the EU AI Act that Al didn't go into detail – things such as :
    - Requirements for product manufacturers, authorized representatives, importers third parties and distributers of Hi-Risk AI Systems
    - Conformity Assessments and issuance of Conformity Certificates
    - Processing of personal data
    - Development, testing and validation of innovative AI systems
    - Establishment of the European Artificial Intelligence Board
    - Establishment of national competent authorities
    - Post-market monitoring by providers and post-market monitoring plan for high-risk AI systems
    - Reporting of serious incidents and of malfunctioning
    - Procedure for dealing with AI systems presenting a risk at national level

4. Al presented the second of the meeting's week's special topics on the NIST AI Risk Management Framework (NIST AI RMF). The reason this second special topic was done was because Al was interested in what the US is doing with respect to AI cybersecurity, and this was one of the things Al found in this area. The slides Al used can be found at https://ftp.pwg.org/pub/pwg/ids/Presentation/NIST AI Risk Management Framework.pdf.

   The main items covered in the presentation were:

- The goals of the NIST AI RMF are:
    - Be risk-based, resource-efficient, pro-innovation, and voluntary
    - Be consensus-driven and developed and regularly updated through an open, transparent process
    - Use clear and plain language that is understandable by a broad audience, including senior executives, government officials, non-governmental organization leadership, and those who are not AI professionals – while still of sufficient technical depth to be useful to practitioners
    - Allow for communication of AI risks across an organization, between organizations, with customers, and to the public at large
    - Provide common language and understanding to manage AI risks
    - Be easily usable and fit well with other aspects of risk management
    - Be useful to a wide range of perspectives, sectors, and technology domain
    - Be outcome-focused and non-prescriptive but not one-size-fits-all requirements
    - Take advantage of and foster greater awareness of existing standards, guidelines, best practices, methodologies, and tools for managing AI risks
    - Be law- and regulation-agnostic
    - Be a living document
    - Offer a resource for improving the ability of organizations to manage AI risks to maximize benefits and to minimize AI-related harms to individuals, groups, organizations, and society

The two that caught Al's eye were (1) the requirement to law and regulation agnostic, which meant that this framework was designed to be independent from addressing any specific NIST, ISO or other standard or from any US or state laws governing risk management, and (2) the requirement that the NIST AI RMF has to be a living document (that is usually assumed but rarely specifically required).

- The current version of the NIST AI RMF is Draft 2 dated August 2022,
- The scope of the NIST AI RMF is:
  - Address challenges unique to AI systems and encourage and equip different AI stakeholders to manage AI risks proactively and purposefully
  - Describes a process for managing AI risks across a wide spectrum of types, applications, and maturity – regardless of sector, size, or level of familiarity with a specific type of technology
  - Intended to be used by AI actors, defined by the Organization for Economic Co-operation and Development (OECD) as "*those who play an active role in the AI system lifecycle, including organizations and individuals that deploy or operate AI*"

  It was noted in the NIST AI RMF that (1) it is a voluntary framework designed to be flexible, (2) it is not a checklist and is not intended to be used in isolation and (3) it is not a compliance mechanism and not intended to supersede existing regulations, laws, or other mandates
- Slide 4 provides a table that shows the AI development lifecycle steps mapped to the AI activities and AI actors which Al just went through very quickly.
- A couple key definitions to understand the NIST AI RMF are:
  - **Risk**: the composite measure of an event's probability of occurring and the magnitude (or degree) of the consequences of the corresponding events
  - **Risk Management**: coordinated activities to direct and control an organization with regard to risk
  - **Risk Tolerance**: The organizations or stakeholder's readiness or appetite to bear the risk in order to achieve its objectives
  - **Reliability:** Ability of an item to perform as required, without failure, for a given time interval, under given conditions
  - **Robustness or generalizability**: Ability of an AI system to maintain its level of performance under a variety of circumstances

  Robustness is something that is hard to quantify and measure, but is an important concept in the NIST AI RMF.
- The NIST AI RMF document listed some key challenge that were faced in developing the framework:
  - AI risks and impacts that are not well-defined or adequately understood are difficult to measure quantitatively or qualitatively
  - Cannot prescribe risk tolerance – need to equip organizations to define reasonable risk tolerance, manage those risks, and document their risk management process
  - Attempting to eliminate risk entirely can be counterproductive in practice – because incidents and failures cannot be eliminated – and may lead to unrealistic expectations and resource allocation that may exacerbate risk and make risk triage impractical
  - Need to integrate AI risks with other critical risks

  The one challenge that stuck out to Al was about attempting to eliminate risks being counterproductive; it was snice to see that clearly stated.
- Probably the main theme of the NIST AI RMF is the concept of "trustworthiness", which in the context of the NIST AI RMF is defined in terms of 7 characteristics - valid and reliable, safe, fair and bias is managed, secure and resilient, accountable and transparent, explainable and

interpretable, and privacy-enhanced, which is pictorially shown in Slide 7. Ira mentioned that ETSI (European Telecommunications Standards Institute) is developing guidelines on “explainability” which is one of the seven Trustworthiness characteristics.

Slide 8 shows a mapping of these 7 characteristics to characteristics in three other related documents, one of which is the EU AI Act which includes the characteristics of robustness and transparency mention in Al’s EU AI Act presentation.

The definitions of these 7 Trustworthiness characteristics are:

- **Valid and Reliable -** should consider that certain types of failures can cause greater harm – and risks should be managed to minimize the negative impact of those failures
- **Safe** - Should not, under defined conditions, cause physical or psychological harm or lead to a state in which human life, health, property, or the environment is endangered
- **Fair and Bias is Managed** - Includes concerns for equality and equity by addressing issues such as bias and discrimination
- **Secure and Resilient** - AI systems that can withstand adversarial attacks, or more generally, unexpected changes in their environment or use, or to maintain their functions and structure in the face of internal and external change, and to degrade gracefully when this is necessary
- **Transparent and Accountable** – Reflects the extent to which information is available to individuals about an AI system, if they are interacting – or even aware that they are interacting – with such a system
- **Explainable and Interpretable** - A representation of the mechanisms underlying an algorithm’s operation, whereas interpretability refers to the meaning of AI systems’ output in the context of its designed functional purpose
- **Privacy-Enhanced** - Norms and practices that help to safeguard human autonomy, identity, and dignity

- The four “Core Steps” of the NIST AI RMF, shown pictorially in Slide 10, are:
  - **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

    Note: Just like the EU AI Act, measurement and metrics is an important aspect of the NIST AI RMF.
  - **Manage**: Entails allocating risk management resources to mapped and measured risks on a regular basis and as defined by the Govern function

  Slides 11-21 listed the categories and subcategories associated with each of the four steps. Al did not go into detail on all the various categories and subcategories, but did discuss a couple categories and/or subcategories for each of the four steps.

  For **Govern**, Al pointed to:

  - Categories “Policies, processes, procedures, and practices across the organization related to the mapping, measuring, and managing of AI risks are in place, transparent, and implemented effectively”, “Processes are in place for robust stakeholder engagement:, and “Policies and procedures are in place to address AI risks arising from third-party software and data and other supply chain issues” as showing the idea of building a Risk Management culture within the organization.
  - The various subcategories like “The characteristics of trustworthy AI are integrated into organizational policies, processes, and procedures” internalize the setting of a RMF culture and the concept of trustworthiness.

- It was nice to see the category "Workforce diversity, equity, inclusion, and accessibility processes are prioritized in the mapping, measuring, and managing of AI risks throughout the lifecycle" included to enforce diversity and equality explicitly.

For **Map**, Al pointed to:

- The category "Context is established and understood" and its associated subcategories like "The organization's mission and relevant goals for the AI technology are understood" and System requirements (e.g., "the system shall respect the privacy of its users") are elicited and understood from stakeholders. Design decisions take socio-technical implications into account to address AI risks" mean that "context" in the NIST AI RMF actually refers to setting up the proper infrastructure , knowledge, awareness and training to implement AI risk management.

For **Measure**, Al pointed to:

- The category "Appropriate methods and metrics are identified and applied", but it will be important for NIST to define what the "appropriate methods and metrics" in the context of the AI RMF are.

- The category "Systems are evaluated for trustworthy characteristics" is important, but like for the first category it will be interesting what metrics NIST proposes to measure each of the subcategories for this category.

For **Manage**, Al pointed to:

- The category "Context is established and understood" and its associated subcategories like "The organization's mission and relevant goals for the AI technology are understood" and System requirements (e.g., "the system shall respect the privacy of its users") are elicited and understood from stakeholders. Design decisions take socio-technical implications into account to address AI risks" mean that "context" in the NIST AI RMF actually refers to setting up the proper infrastructure , knowledge, awareness and training to implement AI risk management.

5. There was a brief discussion of the proposed content for the upcoming IDS Face-to-Face on February 9th. Based on the two special topics presented at today's meeting, Al has changed what he plans to discuss for the special topic at the IDS Face-to-Face; the special topic will compare the approaches to cybersecurity standards and certification between the EU and the US.

   Regarding the status of the HCD iTC, since the HCD cPP v1.0 and HCD SD v1.0 have been published, the focus of that discussion at the IDS Face-to-Face will be on the content and plans for the next (and future) update to the HCD cPP and HCD SD and on the implementation of the HCD Interpretation Team.

6. Round Table:

   - Ira mentioned that the NIST Cybersecurity Framework (CSF) 2.0 | Workshop #2 will be held February 15, 2023, 9:00am - 5:30pm EST. See https://www.nist.gov/news-events/events/2023/02/journey-nist-cybersecurity-framework-csf-20-workshop-2 to register for the event.

7. **Actions:** None

Next Steps

- The IDS WG Session at the PWG February 2023 Face-to-Face Meetings will be on February 9th from 10A – 12N ET.

- The next IDS WG Meeting will be February 23, 2023 at 3:00P ET / 12:00N PT. Main topics will be a TBD special topic (possibly the EU Cybersecurity Act), latest status of the HCD iTC and a post-mortem on the IDS WG Session at the PWG February 2023 Face-to-Face Meetings.