© Peter Cui Bide | petercui.deviantart.com

HISPI Trusted Artificial Intelligence (TAI) Model Top 20

Taiye Lambo

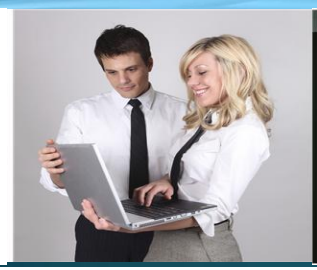
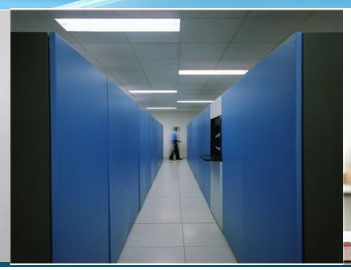
1st and Former CISO, City of Atlanta

Founder, Holistic Information Security Practitioner Institute (HISPI)

October 19, 2023





CYBERSECURITY AWARENESS MONTH 2023

©2023 HISPI

theCYBERIST Outreach

CYBERIST is the Holistic Information Security Practitioner Institute (HISPI) diversity-first outreach program to help strengthen the PEOPLE aspect of Cybersecurity (consisting of People, Processes, and Technology).

@2023 HISPI

Project Cerebellum

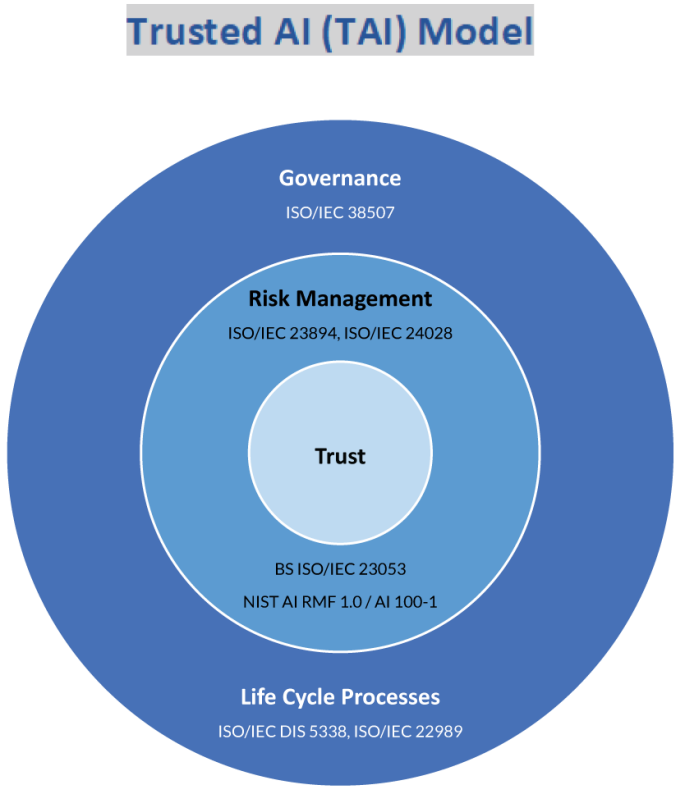
We believe AI should cause no harm,

but enhance the quality of human life.

@2023 HISPI

VISION & MISSION

Our vision is Serving Safe & Secure AI while our mission is to be the brains behind the promotion and harmonization of best practices, standards, and frameworks for AI and related technologies.



@2023 HISPI

Risk of AI

AI can be a powerful tool to reduce the workload of employees, and organizations; but there are risks associated with using AI that need to be mitigated.

On June 29, Australian Research Council announced that they had received applications generated by ChatGPT; prompting them to release a statement advising against it as it “may constitute a breach of confidentiality”.

This is just one example of the risks Generative AI can present to an organization.

@2023 HISPI

HISPI Project Cerebellum Trusted AI Model v1.0

Trusted AI (TAI) Model was created to help guide you through the introduction of AI to your organization and to mitigate the risks associated with its use.

- GOVERN
- MAP
- MEASURE
- MANAGE

@2023 HISPI

Number 1 (Govern 2.2)

Control Group Name	Control Area	Control Specification	Mappings/Crosswalk
GOVERN	Accountability structures are in place so that the appropriate teams and individuals are empowered, responsible, and trained for mapping, measuring, and managing AI risks.	The organization's personnel and partners receive AI risk management training to enable them to perform their duties and responsibilities consistent with related policies, procedures, and agreements.	**ISO/IEC TR 24028:2020** 5.4 **ISO 31000:2018** 5.4.1

PEOPLE	PROCESS	DATA	TECHNOLOGY

@2023 HISPI

Number 2 (Govern 6.1)

Control Group Name	Control Area	Control Specification	Mappings/Crosswalk
GOVERN	Policies and procedures are in place to address AI risks and benefits arising from third-party software and data and other supply chain issues.	Policies and procedures are in place that address AI risks associated with third-party entities, including risks of infringement of a third-party's intellectual property or other rights.	**ISO/IEC TR 24028:2020** 5.5 **ISO 31000:2018** 5.4.1

PEOPLE	PROCESS	DATA	TECHNOLOGY

@2023 HISPI

Number 3 (Map 1.2)

Control Group Name	Control Area	Control Specification	Mappings/Crosswalk
MAP	Context is established and understood.	Interdisciplinary AI actors, competencies, skills, and capacities for establishing context reflect demographic diversity and broad domain and user experience expertise, and their participation is documented. Opportunities for interdisciplinary collaboration are prioritized.	**ISO/IEC TR 24028:2020** 9.8

PEOPLE	PROCESS	DATA

@2023 HISPI

Number 4 (Map 1.6)

Control Group Name	Control Area	Control Specification	Mappings/Crosswalk
MAP	Context is established and understood.	System requirements (e.g., “the system shall respect the privacy of its users”) are elicited from and understood by relevant AI actors. Design decisions take socio-technical implications into account to address AI risks.	**ISO/IEC TR 24028:2020** 9.8 **BS ISO/IEC 23053:2022** 8.3 **ISO 31000:2018** 5.4.1

PEOPLE	PROCESS	DATA	TECHNOLOGY

@2023 HISPI

Number 5 (Map 4.1)

Control Group Name	Control Area	Control Specification	Mappings/Crosswalk
MAP	Risks and benefits are mapped for all components of the AI system including third-party software and data.	Approaches for mapping AI technology and legal risks of its components – including the use of third-party data or software – are in place, followed, and documented, as are risks of infringement of a third party's intellectual property or other rights.	**ISO/IEC TR 24028:2020** 7.1 **ISO 31000:2018** 5.4.1, 5.6

PEOPLE	PROCESS	DATA	TECHNOLOGY

@2023 HISPI

Number 6 (Map 5.1)

Control Group Name	Control Area	Control Specification	Mappings/Crosswalk
MAP	Impacts to individuals, groups, communities, organizations, and society are characterized.	Likelihood and magnitude of each identified impact (both potentially beneficial and harmful) based on expected use, past uses of AI systems in similar contexts, public incident reports, feedback from those external to the team that developed or deployed the AI system, or other data are identified and documented.	**ISO/IEC TR 24028:2020** 9.4.2, 9.10.5 **ISO 31000:2018** 6.3.4 **ISO/IEC 23894:2023** 6.4.2.6, 6.4.3.2, 6.4.3.3

PEOPLE	PROCESS	DATA	TECHNOLOGY

@2023 HISPI

Number 7 (Measure 2.2)

Control Group Name	Control Area	Control Specification	Mappings/Crosswalk
MEASURE	AI systems are evaluated for trustworthy characteristics.	Evaluations involving human subjects meet applicable requirements (including human subject protection) and are representative of the relevant population.	**ISO/IEC TR 24028:2020** 9.10.2.6 **BS ISO/IEC 23053:2022** 8.2 **ISO/IEC 23894:2023** 6.6

PEOPLE	PROCESS	DATA	TECHNOLOGY

Number 8 (Measure 2.5)

Control Group Name	Control Area	Control Specification	Mappings/Crosswalk
MEASURE	AI systems are evaluated for trustworthy characteristics.	The AI system to be deployed is demonstrated to be valid and reliable. Limitations of the generalizability beyond the conditions under which the technology was developed are documented.	**ISO/IEC TR 24028:2020** 9.7, 9.11.2

PEOPLE	PROCESS	DATA	TECHNOLOGY

Number 9 (Measure 2.6)

Control Group Name	Control Area	Control Specification	Mappings/Crosswalk
MEASURE	AI systems are evaluated for trustworthy characteristics.	The AI system is evaluated regularly for safety risks – as identified in the MAP function. The AI system to be deployed is demonstrated to be safe, its residual negative risk does not exceed the risk tolerance, and it can fail safely, particularly if made to operate beyond its knowledge limits. Safety metrics reflect system reliability and robustness, real-time monitoring, and response times for AI system failures.	**ISO/IEC TR 24028:2020** 7.2, 9.9, 9.11.3 **BS ISO/IEC 23053:2022** 8.1 **ISO 31000:2018** 5.6

PEOPLE	PROCESS	DATA	TECHNOLOGY

Number 10 (Measure 2.9)

Control Group Name	Control Area	Control Specification	Mappings/Crosswalk
MEASURE	AI systems are evaluated for trustworthy characteristics.	The AI model is explained, validated, and documented, and AI system output is interpreted within its context – as identified in the MAP function – to inform responsible use and governance.	**ISO/IEC TR 24028:2020** 9.3.4, 9.3.6, 9.10.2 **BS ISO/IEC 23053:2022** 8.1

PEOPLE	PROCESS	DATA	TECHNOLOGY

Number 11 (Measure 2.10)

Control Group Name	Control Area	Control Specification	Mappings/Crosswalk
MEASURE	AI systems are evaluated for trustworthy characteristics.	Privacy risk of the AI system – as identified in the MAP function – is examined and documented.	**ISO/IEC TR 24028:2020** 9.6, 9.10.4 **BS ISO/IEC 23053:2022** 8.1 **ISO 31000:2018** 5.4.1

PEOPLE	PROCESS	DATA	TECHNOLOGY

Number 12 (Measure 2.11)

Control Group Name	Control Area	Control Specification	Mappings/Crosswalk
MEASURE	AI systems are evaluated for trustworthy characteristics.	Fairness and bias – as identified in the MAP function – are evaluated and results are documented.	**ISO/IEC TR 24028:2020** 8.4 **BS ISO/IEC 23053:2022** 6.3, 8.1, 8.3, A.2.2.2 **ISO 31000:2018** 5.4.1

PEOPLE	PROCESS	DATA	TECHNOLOGY

@2023 HISPI

Number 13 (Measure 3.1)

Control Group Name	Control Area	Control Specification	Mappings/Crosswalk
MEASURE	Mechanisms for tracking identified AI risks over time are in place.	Approaches, personnel, and documentation are in place to regularly identify and track existing, unanticipated, and emergent AI risks based on factors such as intended and actual performance in deployed contexts.	**ISO/IEC TR 24028:2020** 9.10.2.5 **BS ISO/IEC 23053:2022** 8.7

PEOPLE	PROCESS	DATA	TECHNOLOGY

Number 14 (Measure 3.3)

Control Group Name	Control Area	Control Specification	Mappings/Crosswalk
MEASURE	Mechanisms for tracking identified AI risks over time are in place.	Feedback processes for end users and impacted communities to report problems and appeal system outcomes are established and integrated into AI system evaluation metrics.	**ISO/IEC TR 24028:2020** 9.4.2 **BS ISO/IEC 23053:2022** 7.6

PEOPLE	PROCESS	DATA	TECHNOLOGY

Number 15 (Measure 4.3)

Control Group Name	Control Area	Control Specification	Mappings/Crosswalk
MEASURE	Feedback about efficacy of measurement is gathered and assessed.	Measurable performance improvements or declines based on consultations with relevant AI actors, including affected communities, and field data about context relevant risks and trustworthiness characteristics are identified and documented.	**BS ISO/IEC 23053:2022** 7.2 **ISO 31000:2018** 5.4.5, 6.2

PEOPLE	PROCESS	DATA	TECHNOLOGY

Number 16 (Manage 2.1)

Control Group Name	Control Area	Control Specification	Mappings/Crosswalk
MANAGE	Strategies to maximize AI benefits and minimize negative impacts are planned, prepared, implemented, documented, and informed by input from relevant AI actors.	Resources required to manage AI risks are taken into account – along with viable non-AI alternative systems, approaches, or methods – to reduce the magnitude or likelihood of potential impacts.	**BS ISO/IEC 23053:2022** 8.1 **ISO 31000:2018** 5.4.3, 5.4.4

PEOPLE	PROCESS	DATA	TECHNOLOGY

Number 17 (Manage 2.3)

Control Group Name	Control Area	Control Specification	Mappings/Crosswalk
MANAGE	Strategies to maximize AI benefits and minimize negative impacts are planned, prepared, implemented, documented, and informed by input from relevant AI actors.	Procedures are followed to respond to and recover from a previously unknown risk when it is identified.	**ISO/IEC 23894:2023** 5.2, 5.3

PEOPLE	PROCESS	DATA	TECHNOLOGY

@2023 HISPI

Number 18 (Manage 2.4)

Control Group Name	Control Area	Control Specification	Mappings/Crosswalk
MANAGE	Strategies to maximize AI benefits and minimize negative impacts are planned, prepared, implemented, documented, and informed by input from relevant AI actors.	Mechanisms are in place and applied, and responsibilities are assigned and understood, to supersede, disengage, or deactivate AI systems that demonstrate performance or outcomes inconsistent with intended use.	**ISO/IEC 23894:2023** 6.4.2.4, 6.4.2.5, 6.4.2.6 **ISO 31000:2018** 6.3.4, 6.5.2

PEOPLE	PROCESS	DATA	TECHNOLOGY

@2023 HISPI

Number 19 (Manage 4.1)

Control Group Name	Control Area	Control Specification	Mappings/Crosswalk
MANAGE	Risk treatments, including response and recovery, and communication plans for the identified and measured AI risks are documented and monitored regularly.	Post-deployment AI system monitoring plans are implemented, including mechanisms for capturing and evaluating input from users and other relevant AI actors, appeal and override, decommissioning, incident response, recovery, and change management.	**BS ISO/IEC 23053:2022** 8.2, 8.6 **ISO/IEC 23894:2023** 5.4.2 **ISO 31000:2018** 6.6

PEOPLE	PROCESS	DATA	TECHNOLOGY

@2023 HISPI

Number 20 (Manage 4.3)

Control Group Name	Control Area	Control Specification	Mappings/Crosswalk
MANAGE	Risk treatments, including response and recovery, and communication plans for the identified and measured AI risks are documented and monitored regularly.	Incidents and errors are communicated to relevant AI actors, including affected communities. Processes for tracking, responding to, and recovering from incidents and errors are followed and documented.	**BS ISO/IEC 23053:2022** 3.1.4, A.2.2.2 **ISO 31000:2018** 6.5.2, 6.7

PEOPLE	PROCESS	DATA	TECHNOLOGY

@2023 HISPI

Questions?

Feel free to connect with me on LinkedIn:
https://www.linkedin.com/in/taiyelambo

E-mail: tlambo@hispi.org

HISPI AI Think Tank:
https://projectcerebellum.com

Personal brand website:
https://taiyelambo.com

@2023 HISPI