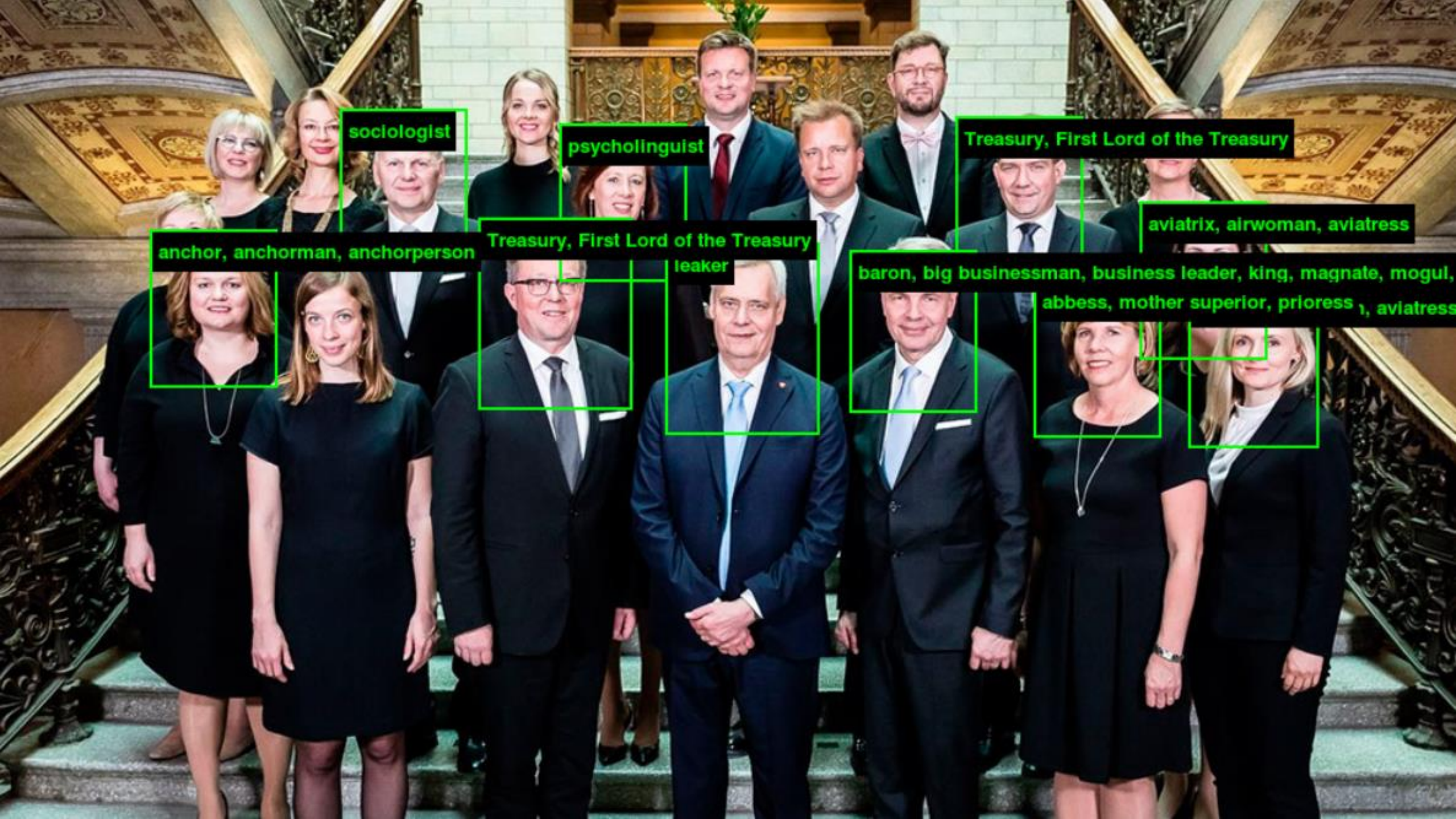
Saidot

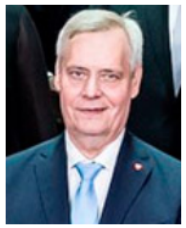
Tekoälyn käyttö ja eettinen pohdinta suomalaisessa yhteiskunnassa

Työterveyshuollon neuvottelukunnan syysseminaari 25.9.2019

Meeri Haataja, CEO & Co-Founder, Saidot – @meerihaataja

saidot.ai


sociologist
psycholinguist
Treasury, First Lord of the Treasury
aviatrix, airwoman, aviatress
anchor, anchorman, anchorperson
Treasury, First Lord of the Treasury
leaker
baron, big businessman, business leader, king, magnate, mogul,
abbess, mother superior, prioress
, aviatress



leaker: *a surreptitious informant*

- person, individual, someone, somebody, mortal, soul > communicator > informant, source > leaker

Treasury, First Lord of the Treasury: *the British cabinet minister responsible for economic strategy*

- person, individual, someone, somebody, mortal, soul > leader > head, chief, top dog > administrator, decision maker > executive, executive director > minister, government minister > cabinet minister > Treasury, First Lord of the Treasury

aviatrix, airwoman, aviatress: *a woman aviator*

- person, individual, someone, somebody, mortal, soul > worker > skilled worker, trained worker, skilled workman > aviator, aeronaut, airman, flier, flyer > aviatrix, airwoman, aviatress

sociologist: *a social scientist who studies the institutions and development of human society*

- person, individual, someone, somebody, mortal, soul > scientist > social scientist > sociologist

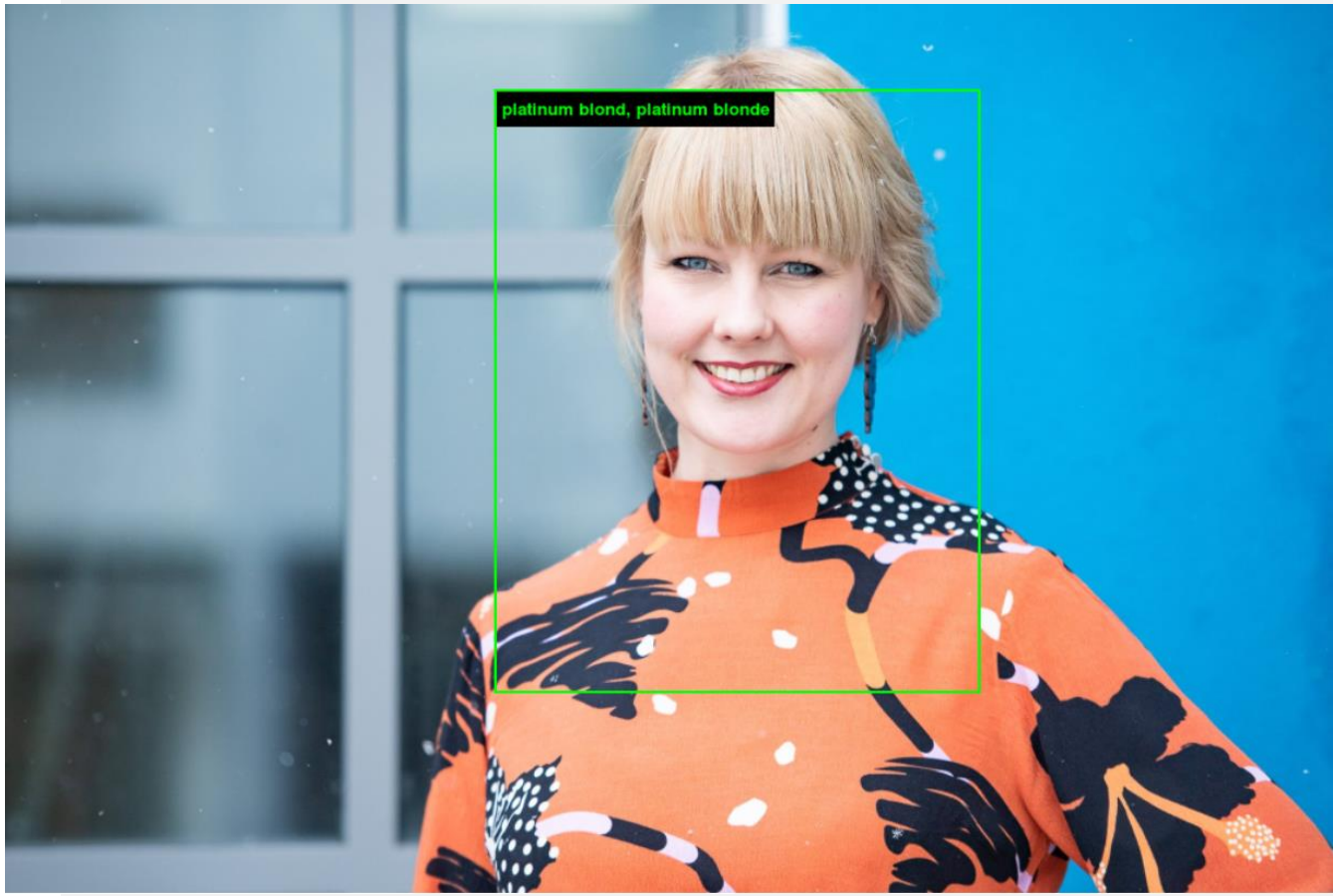
baron, big businessman, business leader, king, magnate, mogul, power, top executive, tycoon: *a very wealthy or powerful businessman*

- person, individual, someone, somebody, mortal, soul > capitalist > businessperson, bourgeois > businessman, man of affairs > baron, big businessman, business leader, king, magnate, mogul, power, top executive, tycoon

Meeri Haataja

CEO & Co-Founder
+358407725010
meeri@saidot.ai
@meerihaataja

- Over 18 years of experience on driving data, analytics and AI use as well as privacy in large enterprises in financial services, telecommunications, high-tech and media.
- Chair of IEEE's Ethics Certification Program for Autonomous & Intelligent Systems.
- Affiliate at the Berkman Klein Center for Internet & Society at Harvard University (2019-2020).
- Previous chair of Ethics working group of Finland's AI Program.



Saidot

Automated decision systems can exist in any context where government bodies or agencies evaluate people or cases, allocate scarce resources, focus scrutiny or surveillance on communities, or make nearly any sort of decision.

https://ainowinstitute.org/aap-toolkit.pdf

Saidot

Human Resources/Public Benefits

Child Risk and Safety Assessment	An instrument that assesses the risk of current and future harm to a child. The tool can be used at different stages in the decision making process at a child welfare agency. Common uses include workers assessing whether a family should receive a secondary visit by a social service worker, or whether a family should receive intervention services.
Genogram and Ecomap Software	An assessment tool that allows child welfare caseworkers to map family trees, identify gaps in family history, organize information amassed from family, and assess interventions.
Medicaid eligibility assessment	A tool that determines eligibility and compliance for Medicaid. Similar tools are used to assess eligibility and compliance for other public benefits. Known Vendor: **IBM, APS Healthcare**

Criminal Justice

Surveillance Technologies	Many surveillance technologies used by local and state law enforcement use algorithms including but not limited to, facial recognition (including on body cameras), automatic license plate readers, and visual or data analytics systems. Law enforcement agencies also data-mining software that processes large quantities of data from commercial and government sources to identify relationships or connections between people, places, and things. Known Vendors: **Palantir, Vigilant Solutions, Cognitec, Amazon, Microsoft, Motorola, IBM, Axon**
Predictive Policing	Any system that analyzes available data to predict either (A) where a crime may happen in a given time window (place-based) or (B) who will be involved in a crime as either victim or perpetrator (person-based). A predictive policing system then must convey that information to police officers or other social service providers so that they can take some course of action. Known Vendors: **Predpol; Azavea (Hunchlab); Palantir, Starlight, Bair Analytics, IBM, RTMDx**

Education

School Assignment	Many school assignments are determined using a simple match algorithm that evaluates school choices selected by parents and a school districts admission preferences and seat availability. Knowns Vendors: **Institute for Innovation in Public School Choice**[8]
Controlled Choice	A student assignment algorithm that is designed to achieve school diversity and optimize choice/distribution balance, particularly in school districts experiencing racial and/or socioeconomic segregation. Parents rank-order schools by preference, and students are assigned to school based on constraints set by the school district to achieve a balanced student distribution goal. Known Vendors: **Michael Alves (education consultant)**
School Violence Risk Assessment	A tool designed to identify students who are at a high risk for school related violence (e.g. homicide, suicide). A recent study used the BRACHA (Brief Rating of Aggression by Children and Adolescents) scale measures aggressive behavior, and the School Safety Scale, which measures behavioral changes that may indicate violence, and manual annotation of student interview to make predictions about likelihood of violence.

Public Health

Disease treatment	An algorithm used to identify individual with chronic hepatitis C for treatment and cure. The systems also analyzing health surveillance data to monitor treatment and cure rates within a municipality to assess progress towards treatment goals.
Prescription	Some states are using proprietary algorithms applied to prescription drug monitoring

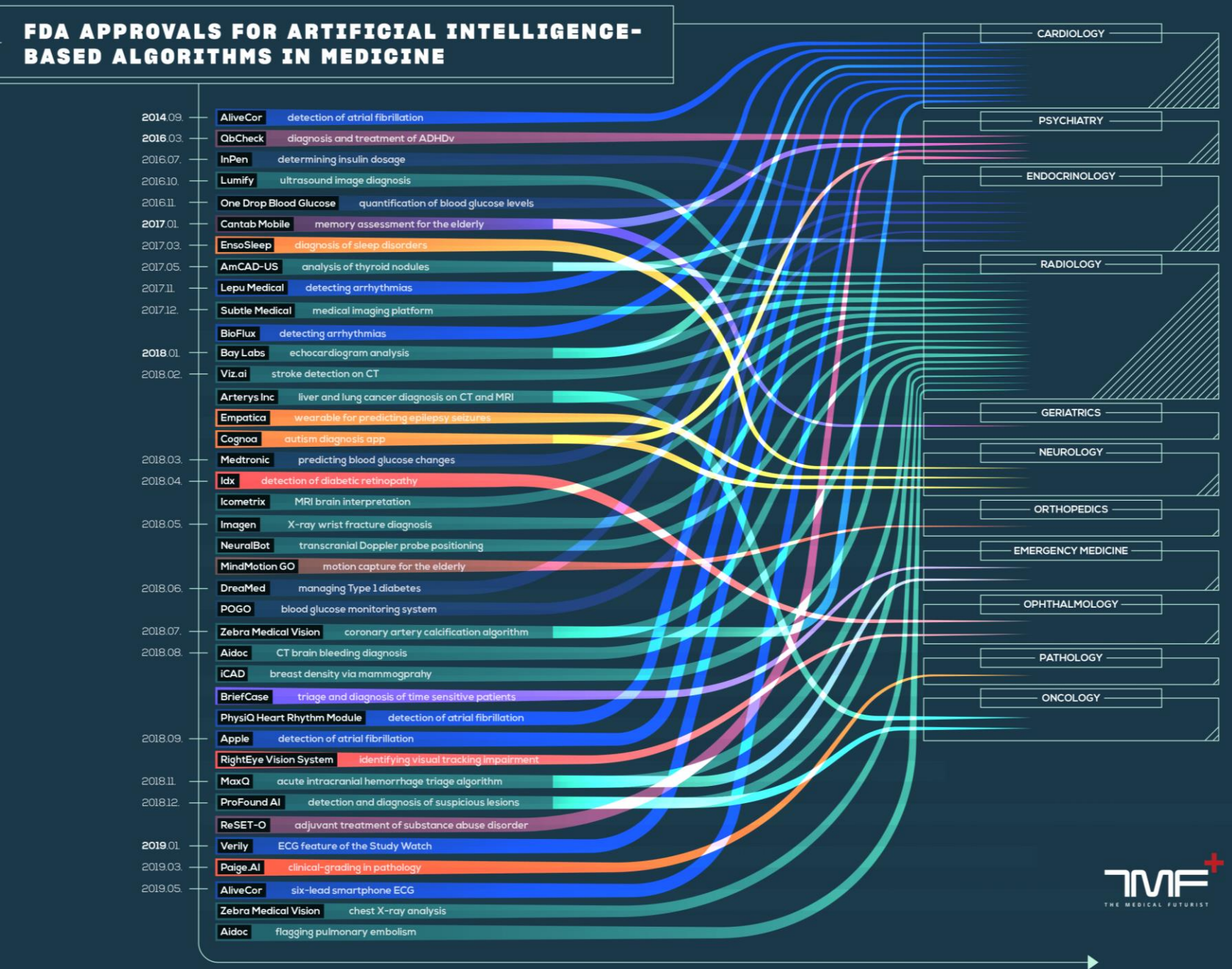
VM Robosaatiohaut 2018

Hyväksyttynä 39 hanketta lähes kahdessakymmenessä virastossa

- ✚ Etelä-Savon ELY-keskus: KasvuBot – älykäs kasvupalveluohjaaja
- ✚ Kilpailu- ja kuluttajavirasto: Koneoppimisen hyödyntäminen KUTI-järjestelmän tiedon laadun parantamisessa sekä automaattiset seurantatyökalut KUTI-järjestelmässä
- ✚ Kilpailu- ja kuluttajavirasto: Kartellitutka
- ✚ Luonnonvarakeskus: Hyönteistuholaisten seuranta -pilotti
- ✚ Maahanmuuttovirasto: PuheKamu
- ✚ Maahanmuuttovirasto: Tiedon jalostus keinoälyllä
- ✚ Merivoimien esikunta: VR/AR/MR-teknologian tutkimus ja implementointi PASSI-järjestelmään (VEERA)
- ✚ Oikeusministeriö: Kokeiluhanke liittyen sähköisen lausuntomenettelyn kautta tulleiden lausuntojen analysointiin ja raportointiin luonnollisen kielen analyysin keinoin
- ✚ Oikeusrekisterikeskus: Oikeusrekisterikeskuksen kansalais- ja viranomaispalvelujen maksujen käsittelyrobotti
- ✚ Puolustusvoimat, Pääesikunta, Suunnitteluosasto: Mobiili riski- ja poikkeamailmoitus ”Lassie”
- ✚ Puolustusvoimat, Pääesikunta: HATKAT - Hallinnollisen taakan vähentäminen
- ✚ Puolustusvoimat: PVRobo
- ✚ Ruokavirasto: AI Elintarvikevalvonnan dataan
- ✚ Ruokavirasto: Puheentunnistus lihantarkastukseen
- ✚ Ruokavirasto: Robotiikan kehittäminen Ruokaviraston asiakaspalveluun
- ✚ Ruokavirasto: Siemenpäätösten automatisointi
- ✚ Ruokavirasto: Tekoälyassistentti tiedonhaussa
- ✚ Sosiaali- ja terveysministeriö: Ensi- ja akuuttihoidon hoidon tarpeen arvion vaikuttavuus potilasturvallisuuteen
- ✚ Sosiaali- ja terveysministeriö: Tekoälyn hyödyntäminen terveyteen, hyvinvointiin ja turvallisuuteen liittyvien ilmiöiden ennakoinnissa ja tunnistamisessa
- ✚ Tulli: Koneoppimisen esitutkimus, mallinnus ja prototyypin rakentaminen
- ✚ Turvallisuus- ja kemikaalivirasto: NettiDogi
- ✚ Valtiokonttori: Asennusautomaatioteknologian pilotointi kapasiteettipalvelussa
- ✚ Valtiokonttori: Lohkoketjuteknologian hyödyntäminen liikennevakuutuksen laiminlyöntimaksujen käsittelyyn
- ✚ Valtion talous- ja henkilöstöhallinnon palvelukeskus Palkeet: BI Digicontroller
- ✚ Valtion talous- ja henkilöstöhallinnon palvelukeskus Palkeet: Palkkionsaajien sähköinen asiointi (ePalkkio)
- ✚ Valtion tieto- ja viestintätekniikkakeskus Valtori: ContactCenter (OC-SaaS) järjestelmien automaatti raportointi ja OC-järjestelmien käyttövaltuushallinta
- ✚ Valtion tieto- ja viestintätekniikkakeskus Valtori: Massaviestinnän toimintatavan automatisointi
- ✚ Valtion tieto- ja viestintätekniikkakeskus Valtori: Säännöllisten vakioraporttien tuotannon automatisointi viidelle palvelulle
- ✚ Valtion tieto- ja viestintätekniikkakeskus Valtori: Työajan kirjaus TOP-toiminnanohjaus-järjestelmästä Kiekuun automaattisesti
- ✚ Valtion tieto- ja viestintätekniikkakeskus Valtori: Virtu IdP2 raportoinnin ja lokianalytiikan automatisointi
- ✚ Valtiovarainministeriö, kansantalousosasto: Dynaamisten raporttien mallipohjien rakentaminen

Saidot

FDA APPROVALS FOR ARTIFICIAL INTELLIGENCE-BASED ALGORITHMS IN MEDICINE



- AliveCor supports the early detection of atrial fibrillation, developed an ECG analytics platform – just as PhysiQ Heart Rhythm Module, Apple, and Cardiologs – and a six-lead smartphone ECG.
- QbCheck helps with the diagnosis and treatment of ADHD.
- InPen tracks insulin dosage.
- One Drop Blood Glucose quantifies blood glucose levels and automatically sends the data to the paired app.
- Lumify offers ultrasound image diagnosis.
- Cantab Mobile acts as a tool for memory problem assessment for the elderly.
- EnsoSleep powers a tool for recognizing sleep disorders.
- AmCAD-US evaluates thyroid nodules and categorizes nodule characteristics.
- Lepu Medical and BioFlux detect arrhythmias.
- Subtle Medical offers a medical imaging platform.
- Bay Labs offers echocardiogram analysis.
- Viz.AI detects stroke on CT scans and helps clinicians win the race against time.
- Arterys' algorithm is able to spot cancerous lesions in liver and lungs on CT and MR images.
- Empatica helps predict epileptic seizures.
- Cognoa's algorithm built into an app helps diagnose autism in kids.
- Medtronic and POGO monitor and predict blood glucose changes.
- Idx autonomously detects diabetic retinopathy using retinal images.
- Icometrix helps neurologists interpret brain MR images.
- Imagen aids healthcare providers in identifying wrist fractures with similar accuracy as human radiologists.
- NeuralBot offers a solution for transcranial Doppler probe positioning.
- MindMotion Go advances its algorithm for motion capture for the elderly.
- Dreamed assists healthcare professionals in the management of Type 1 diabetes.
- Zebra Medical Vision detects, quantifies coronary artery calcification, and analyses chest X-rays.
- Aidoc is able to flag brain bleeding in head CT images and pulmonary embolism.
- iCAD classifies breast density and detects breast cancer as accurately as radiologists.
- ScreenPoint Medical assists radiologists with the reading of screening mammograms.
- Briefcase triages and diagnoses time-sensitive patients.
- RightEye Vision System tracks eye movements for identifying visual tracking impairment.
- MaxQ develops an acute intracranial hemorrhage triage algorithm.
- ProFound AI detects and diagnoses suspicious lesions.
- ReSET-O offers an adjuvant treatment of substance abuse disorder.
- Verily developed an ECG feature on the Study Watch.
- Paige.AI provides a clinical-grade algorithm in pathology.
- FerriSmart created a machine learning solution for the quantification of liver iron concentration.

Loss of privacy

Digital manipulation

Algorithmic bias & inequality

Marginalizing the marginalized

Social scoring & mass surveillance

Machines as morale agents

Adversarial AI & data thefts

Loss of accountability

Loss of work

Saidot

Mistaken AI
Malicious AI
Adversarial AI

Some key concerns on public use of AI

Fairness and non-discrimination in use of public power

Protection of vulnerable groups

Autonomy and freedoms

Safety and security

...in high stake applications like...

Identifying and tracking individuals

AI enabled citizen scoring

Judicial applications

Public health care & education

Autonomous weapon systems

Saidot

Perustuslaki

6 § Yhdenvertaisuus
Ihmiset ovat yhdenvertaisia lain edessä. Ketään ei saa ilman hyväksyttävää perustetta asettaa eri asemaan sukupuolen, iän, alkuperän, kielen, uskonnon, vakaumuksen, mielipiteen, terveydentilan, vammaisuuden tai muun henkilöön liittyvän syyn perusteella. Lapsia on kohdeltava tasa-arvoisesti yksilöinä, ja heidän tulee saada vaikuttaa itseään koskeviin asioihin kehitystään vastaavasti. Sukupuolten tasa-arvoa edistetään yhteiskunnallisessa toiminnassa sekä työelämässä, erityisesti palkkauksesta ja muista palvelussuhteen ehdoista määrättäessä, sen mukaan kuin lailla tarkemmin säädetään.

7 § Oikeus elämään sekä henkilökohtaiseen vapauteen ja koskemattomuuteen
Jokaisella on oikeus elämään sekä henkilökohtaiseen vapauteen, koskemattomuuteen ja turvallisuuteen. Ketään ei saa tuomita kuolemaan, kiduttaa eikä muutoinkaan kohdella ihmisarvoa loukkaavasti. Henkilökohtaiseen koskemattomuuteen ei saa puuttua eikä vapautta riistää mielivaltaisesti eikä ilman laissa säädettyä perustetta. Rangaistuksen, joka sisältää vapaudenmenetyksen, määrää tuomioistuin. Muun vapaudenmenetyksen laillisuus voidaan saattaa tuomioistuimen tutkittavaksi. Vapautensa menettäneen oikeudet turvataan lailla.

Yhdenvertaisuuslaki

8 § Syrjinnän kielto Ketään ei saa syrjiä iän, alkuperän, kansalaisuuden, kielen, uskonnon, vakaumuksen, mielipiteen, poliittisen toiminnan, ammattiyhdistystoiminnan, perhesuhteiden, terveydentilan, vammaisuuden, seksuaalisen suuntautumisen tai muun henkilöön liittyvän syyn perusteella. Syrjintä on kielletty riippumatta siitä, perustuuko se henkilöä itseään vai jotakuta toista koskevaan tosiseikkaan tai oletukseen.

Assigning liability: Responsibility of medical errors

1. Coders and designers
2. Medical device companies
3. Physicians and other health care professionals
4. Hospitals and health care systems
5. Other actors, including regulators, insurance companies, pharmaceutical companies, and medical schools

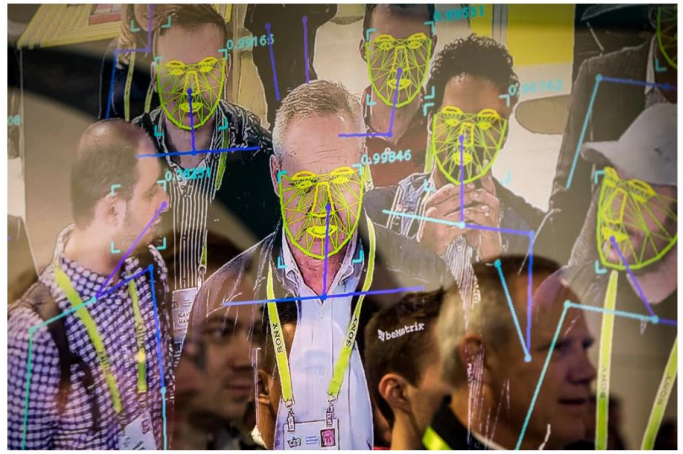
Source: How Should Clinicians Communicate With Patients About the Roles of Artificially Intelligent Team Members? Schiff and Borenstein, AMA Journal of Ethics, February 2019, Vol 21, No 2: E138-145

Saidot

Lack of trust and current trend in AI governance

The New York Times

San Francisco Bans Facial Recognition Technology



Attendees interacting with a facial recognition demonstration at this year's CES in Las Vegas. Joe Buglewicz for The New York Times

By Kate Conger, Richard Fausset and Serge F. Kovaleski

May 14, 2019

366

Somerville Bans Government Use Of Facial Recognition Tech

June 28, 2019 | By Katie Lannan, State House News Service

A security camera in the Financial District of San Francisco (Eric Risberg/AP)

Portland officials want to ban private use of facial recognition technology, citing 'accuracy problems'

BY KATE KAYE on September 5, 2019 at 10:49 am

1 Comment | Share 257 | Tweet | Share | Reddit | Email

Portland City Commissioner Jo Ann Hardesty wants to adopt what could be the most far-reaching ban on facial recognition technology in the country.

Later this month, the city council is expected to evaluate a facial recognition ban proposal which could prevent government agencies from using the identification technology — but that's not all.

Unlike other city-wide bans that stop at government use, Hardesty supports restrictions that would make the controversial technology off-limits to Portland businesses such as retailers using it to discourage would-be thieves from entering their stores, or corporations using it for employee identification and surveillance.

Portland City Commissioner Jo Ann Hardesty. (City of Portland Photo)

Saidot

Making Policy on Augmented Intelligence in Health Care

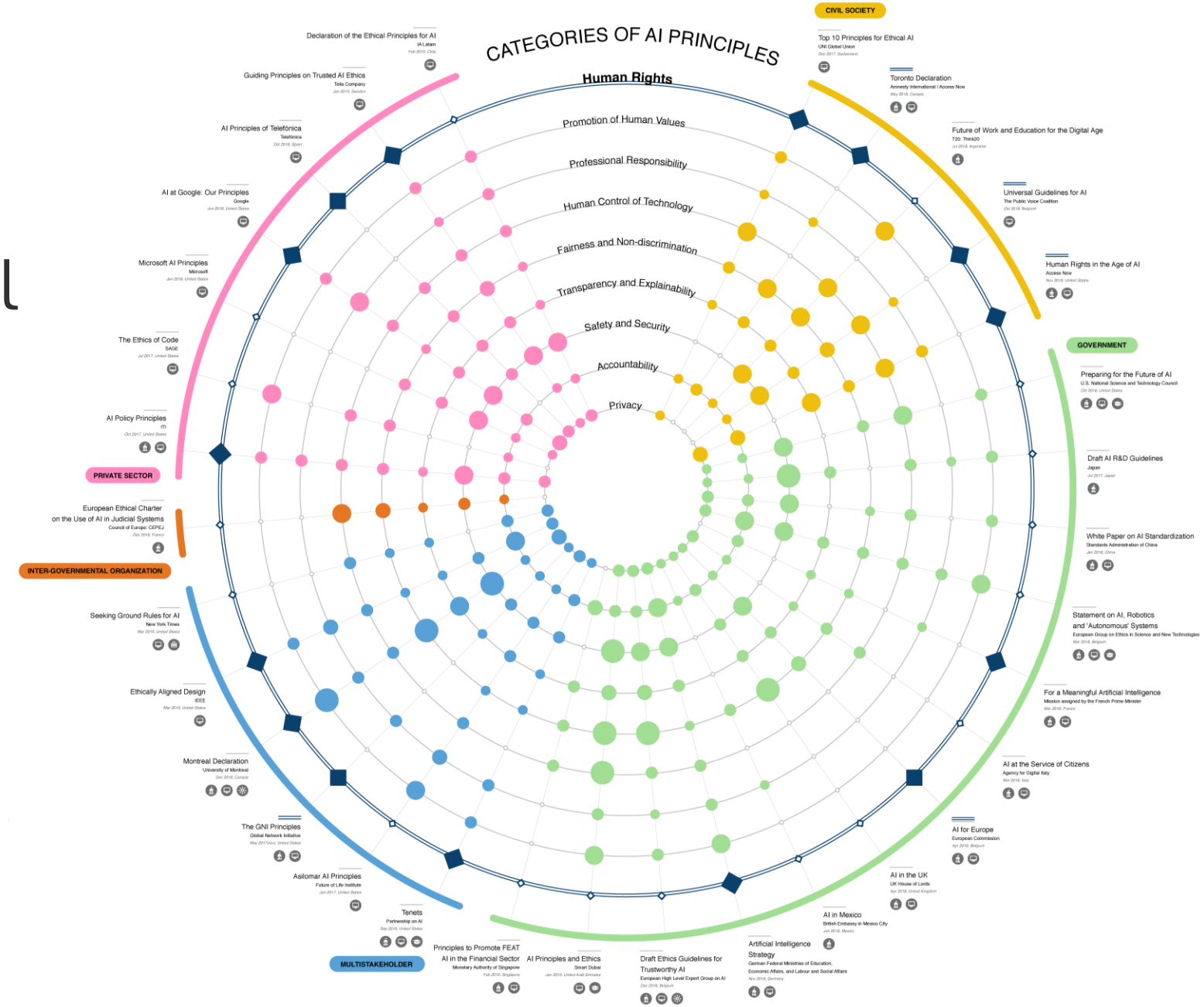
“Health AI must be deployed in ways that promote quality of care and minimize potentially disruptive effects.”

Elliot Crigger & Christopher Khoury on AMA Journal of Ethics

Principled Artificial Intelligence

Berkman Klein's Cyberlaw Clinic launched the "Principles Artificial Intelligence Project" to map AI principles and guidelines. The team created a data visualization to summarize their findings, and will later publish the final data visualization, along with the dataset itself and a white paper detailing their assumptions, methodology and key findings.

https://ai-hr.cyber.harvard.edu/primp-viz.html



Saidot

Guidelines

EU Commission Ethics Guidelines
4 ethical principles

The principle of respect for human autonomy

Humans interacting with AI systems must be able to keep full and effective self-determination over themselves, and be able to partake in the democratic process. AI systems should not unjustifiably subordinate, coerce, deceive, manipulate, condition or herd humans.

The principle of prevention of harm

AI systems should neither cause nor exacerbate harm or otherwise adversely affect human beings. This entails the protection of human dignity as well as mental and physical integrity.

The principle of fairness

The development, deployment and use of AI systems must be fair. Ensuring equal and just distribution of both benefits and costs, and ensuring that individuals and groups are free from unfair bias, discrimination and stigmatisation. Ability to contest and seek effective redress against decisions made by AI systems and by the humans operating them.

The principle of explicability

Explicability is crucial for building and maintaining users' trust in AI systems. This means that processes need to be transparent, the capabilities and purpose of AI systems openly communicated, and decisions – to the extent possible – explainable to those directly and indirectly affected.

Guidelines

EU Commission Ethics Guidelines 7 requirements

1. **Human agency and oversight**

Including fundamental rights, human agency and human oversight

2. **Technical robustness and safety**

Including resilience to attack and security, fall back plan and general safety, accuracy, reliability and reproducibility

3. **Privacy and data governance**

Including respect for privacy, quality and integrity of data, and access to data

4. **Transparency**

Including traceability, explainability and communication

5. **Diversity, non-discrimination and fairness**

Including the avoidance of unfair bias, accessibility and universal design, and stakeholder participation

6. **Societal and environmental wellbeing**

Including sustainability and environmental friendliness, social impact, society and democracy

7. **Accountability**

Including auditability, minimisation and reporting of negative impact, trade-offs and redress

Saidot

High-quality, clinically validated health care AI

1. is designed and evaluated in keeping with best practices in **user-centered design**, particularly for physicians and other members of the health care team;
2. is **transparent;**
3. conforms to leading standards for **reproducibility**;
4. identifies and takes steps to **address bias** and **avoids** introducing or exacerbating **health care disparities** including when testing or deploying new AI tools on vulnerable populations; and
5. safeguards patients' and other individuals' **privacy** interests and preserves the **security** and **integrity of personal information**.

Source: Augmented Intelligence in Health Care, H-480.940 by American Medical Association, 2018

A new enabling principle for AI → Explicability

Synthesis of 47 AI ethics principles

Traditional bioethics principles

1. **Beneficence**
2. **Non-maleficence**
3. **Autonomy**
4. **Justice**

A new enabling principle for AI

5. **Explicability**

Source: Floridi, L., Cowls, J., Beltrametti, M. et al. Minds & Machines (2018) 28: 689.

Saidot

From principles to practice

Transparency: moving from principles to practice

Saidot

The way of transparency

Transparency concerns the **reduction of information asymmetry**.

Transparency is key to **building and maintaining citizen's trust** in the developers of AI systems and AI systems themselves.

— Uttermost critical in question of exercising of the **public power**.

Both **technological and business model** transparency matter from an ethical standpoint.

Source: EU Commission's High level Expert Group on AI

Case for transparency as a fundamental principle

Without transparency we have no means to monitor any other ethics principles, or how the public authority is exercised through algorithmic decisions

Saidot

Perustuslaki

2 § Kansanvaltaisuus ja oikeusvaltioperiaate

…

Julkisen vallan käytön tulee perustua lakiin. Kaikessa julkisessa toiminnassa on noudatettava tarkoin lakia.

12 § Sananvapaus ja julkisuus

…

Viranomaisen hallussa olevat asiakirjat ja muut tallenteet ovat julkisia, jollei niiden julkisuutta ole välttämättömien syiden vuoksi lailla erikseen rajoitettu. Jokaisella on oikeus saada tieto julkisesta asiakirjasta ja tallenteesta.

Julkisuuslaki

1 § Julkisuusperiaate Viranomaisten asiakirjat ovat julkisia, jollei tässä tai muussa laissa erikseen toisin säädetä. Oikeudesta seurata eduskunnan täysistuntoa, valtuuston ja muiden kunnallisten toimielinten kokouksia sekä tuomioistuinten ja kirkollisten toimielinten istuntoja säädetään erikseen.

2 § Lain soveltamisala Tässä laissa säädetään oikeudesta saada tieto viranomaisten julkisista asiakirjoista sekä viranomaisessa toimivan vaitiolovelvollisuudesta, asiakirjojen salassapidosta ja muista tietojen saantia koskevista yleisten ja yksityisten etujen suojaamiseksi välttämättömistä rajoituksista samoin kuin viranomaisten velvollisuuksista tämän lain tarkoituksen toteuttamiseksi. Viranomaisen asiakirjojen tiedonhallinnasta säädetään julkisen hallinnon tiedonhallinnasta annetussa laissa (906/2019). (9.8.2019/907)

3 § Lain tarkoitus Tässä laissa säädettyjen tiedonsaantioikeuksien ja viranomaisten velvollisuuksien tarkoituksena on toteuttaa avoimuutta viranomaisten toiminnassa sekä antaa yksilöille ja yhteisöille mahdollisuus valvoa julkisen vallan ja julkisten varojen käyttöä, muodostaa vapaasti mielipiteensä sekä vaikuttaa julkisen vallan käyttöön ja valvoa oikeuksiaan ja etujaan.

Call for standardized approach

We need a shared standard for algorithmic transparency in public sector to enable citizen agency and allow efficiency in implementation and oversight.

Saidot

Transparency

Three examples on concrete actions

Saidot

Citizen Trust Through AI Transparency

Six Finnish organizations have kicked off a project to standardize public organizations' transparency and communications of the use of AI to citizens. Project will provide **internationally adoptable guidelines on how to inform about public sector use of personal data and AI to the citizens**.

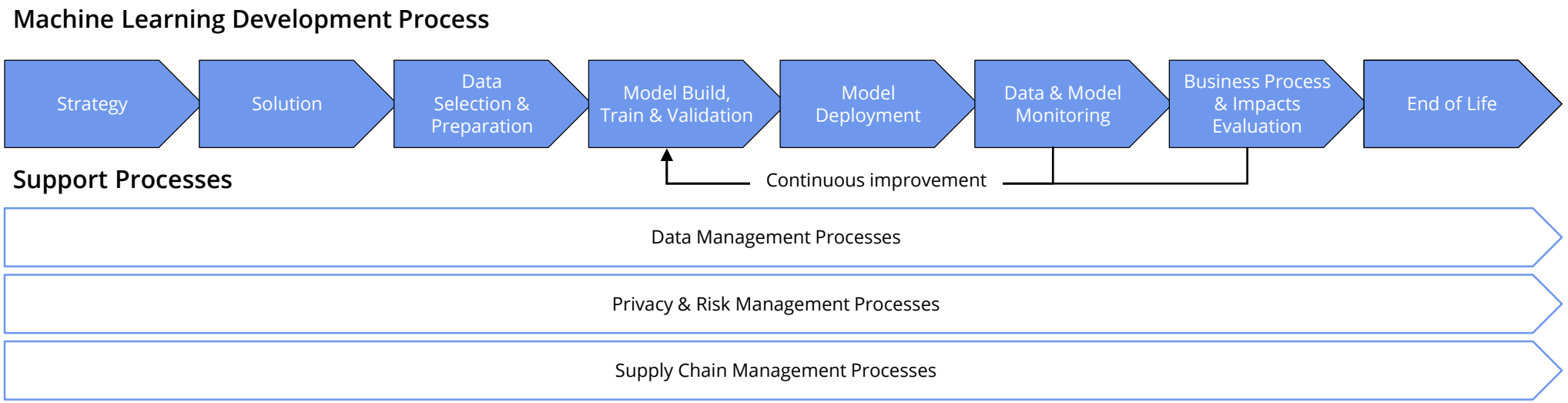
The project participants are **City of Espoo, City of Helsinki, Kela, Finland's Ministry of Justice, The Finnish Innovation Fund Sitra and Saidot**.

They will together

- define what information (metadata) citizens have a right to know of the Finnish public sector AI
- create guidelines on how this information should be communicated to the citizens.

Saidot

Transparency must address the whole lifecycle of an algorithmic system



Saidot

Synthetizing legal, ethical, technical and citizen perspectives into shared approach for public authority algorithmic transparency.

Saidot

Stakeholders need varying levels of transparency

Audience	***User technical skill level***
1. Information for proactive citizen communication	Low
2. Information for human operators 3. Additional information on citizen request	Medium / High
4. Full disclosure of information for independent audits & compliance	High

Algorithmic transparency for public authorities

Preliminary categories for algorithmic transparency & impact assessment

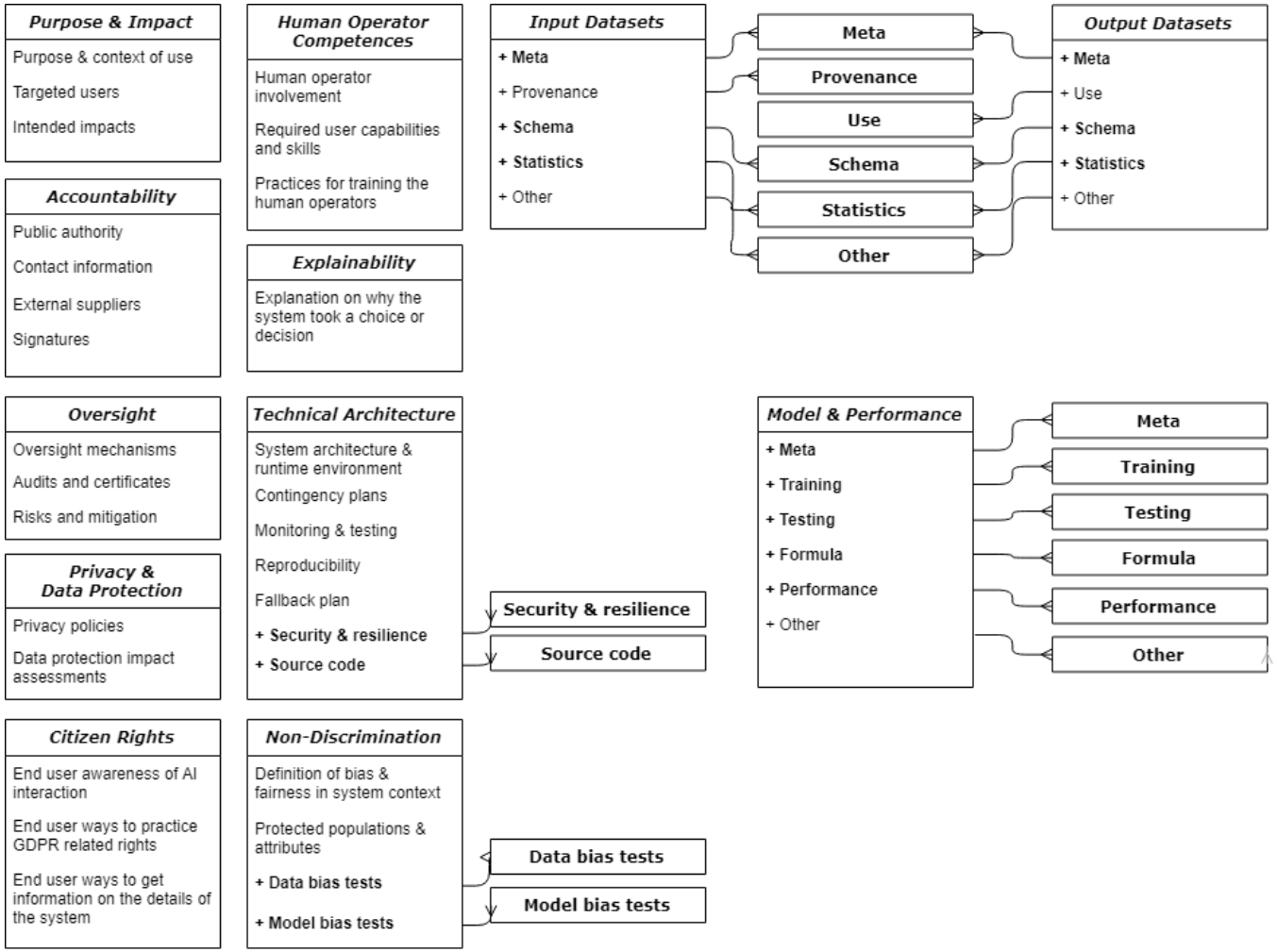
Purpose & Impact	***Accountability***	***Oversight***	***System Behaviour: Technical Architecture***	***System Behaviour: Input Datasets***	***System Behaviour: Model & Performance***
Allows citizens understand why the system exists, how is it used and what kinds of impacts it has in people and wider society.	Reveals who owns the system and who are the accountable organization and civil servants in case of problems or need for further information.	Provides information on whether independent third parties have reviewed, tested or audited the system.	Provides technical information on the wider system context and key integrations, and the processes for safety and security.	Provides information about the data collection and variables a system uses to produce an output.	Provides technical information on how the system processes data to produce its outputs and the accuracy of the predictions.
System Behaviour: Output Datasets	***Non-Discrimination & Fairness***	***Explainability***	***Human Operator Role & Competences***	***Citizen Rights***	***Privacy & Data Protection***
Provides information of the specific outputs of the algorithmic system and the accesses to and uses of this information.	Provides information on how the system addresses non-discrimination and what are the results of such bias tests.	Provides answer to question why the algorithmic system in question makes one prediction and not another.	Provides understanding on the human operator involvement and trainings and instructions given.	Provides information on how to gain more information on the system, and how to exercise GDPR related rights.	Provides information on the relevant privacy policies and possible privacy impact assessments conducted.

Saidot

Algorithmic transparency for public authorities

Preliminary metadata model for open consultation

Category	***Description***
Level of Automatization	Classification of the system into automated decision-making systems, and expert support systems.
Level of Impacts	Possibility to categorize systems based on the scale of impacts and adapt the transparency requirements based on category.



Saidot

Regulating AI

FDA's draft regulative framework for AI

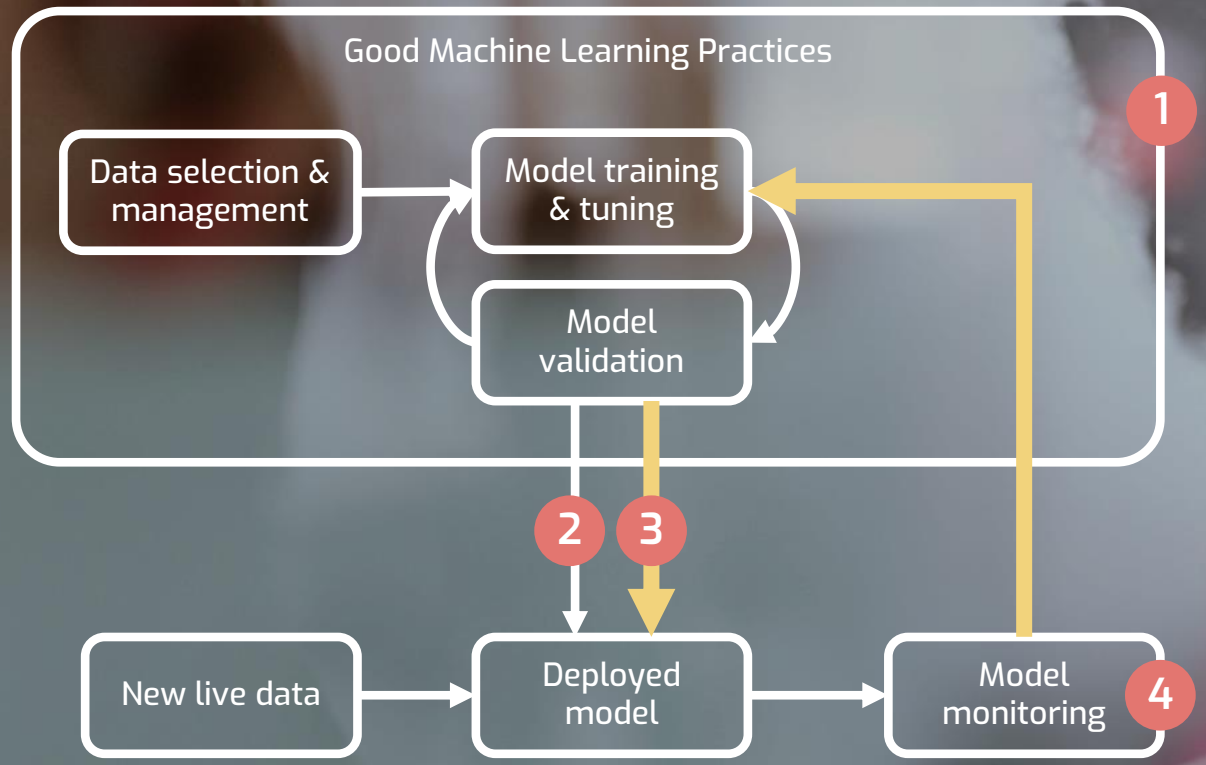
Total product lifecycle approach for adaptive and learning software medical devices assures that ongoing algorithm changes

- follow a pre-specified performance objectives and change control plans,
- use a validation process that ensures improvements to the performance, safety and effectiveness of the artificial intelligence software, and
- includes real-world monitoring of performance once the device is on the market to ensure safety and effectiveness are maintained.

Saidot

Regulating AI

Sneak peak to future regulation?



Overlay of FDA's TPLC approach on AI/ML workflow

FDA's total product life-cycle approach to enable beneficial and innovative AI software to come to market while ensuring the device's benefits continue to outweigh risks

1. **Culture of quality** and organizational excellence
2. **Premarket assurance** of safety and effectiveness
3. **Review** of Software as a Medical Device (SaMD) **pre-specifications** and **algorithm change protocol**
4. Real-world **performance monitoring and evaluation**

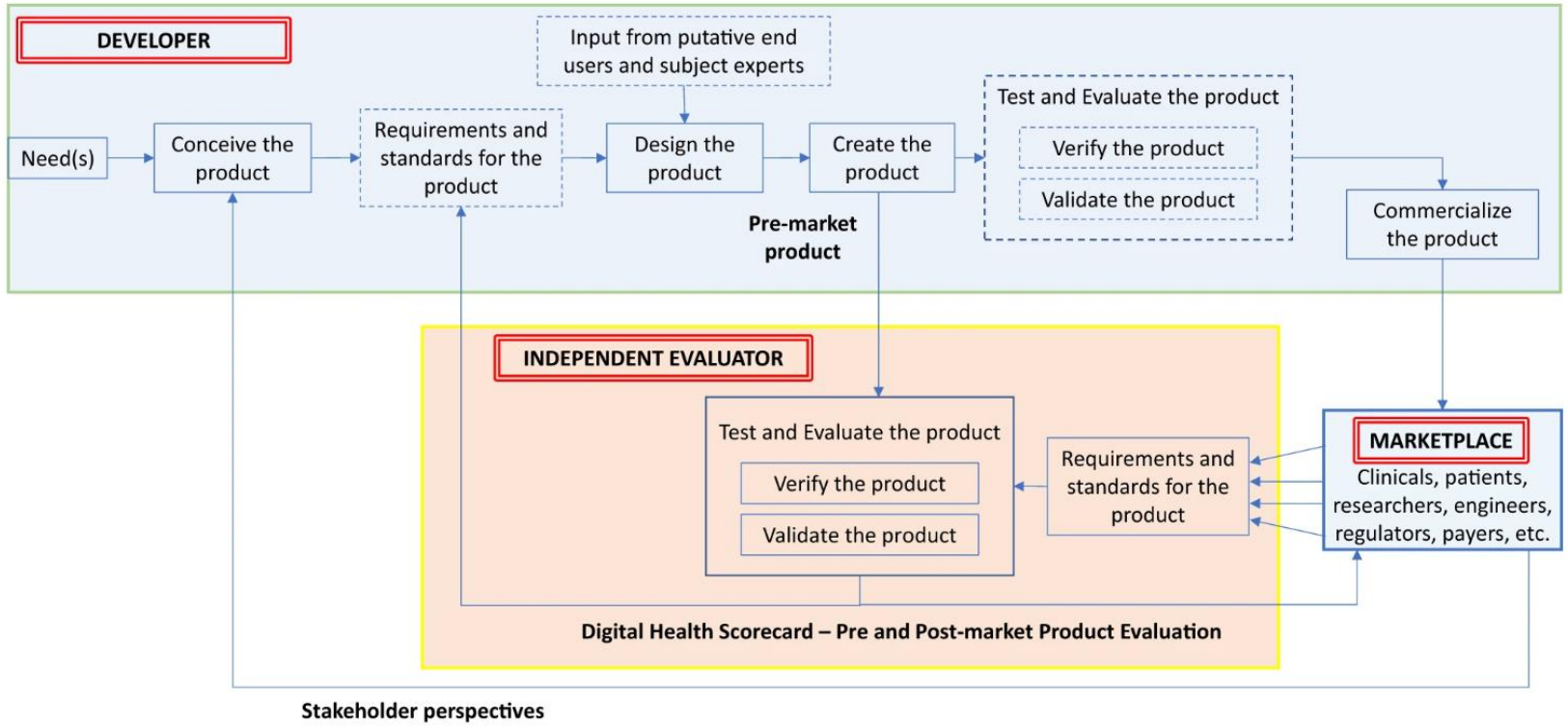
Source: FDA

Saidot

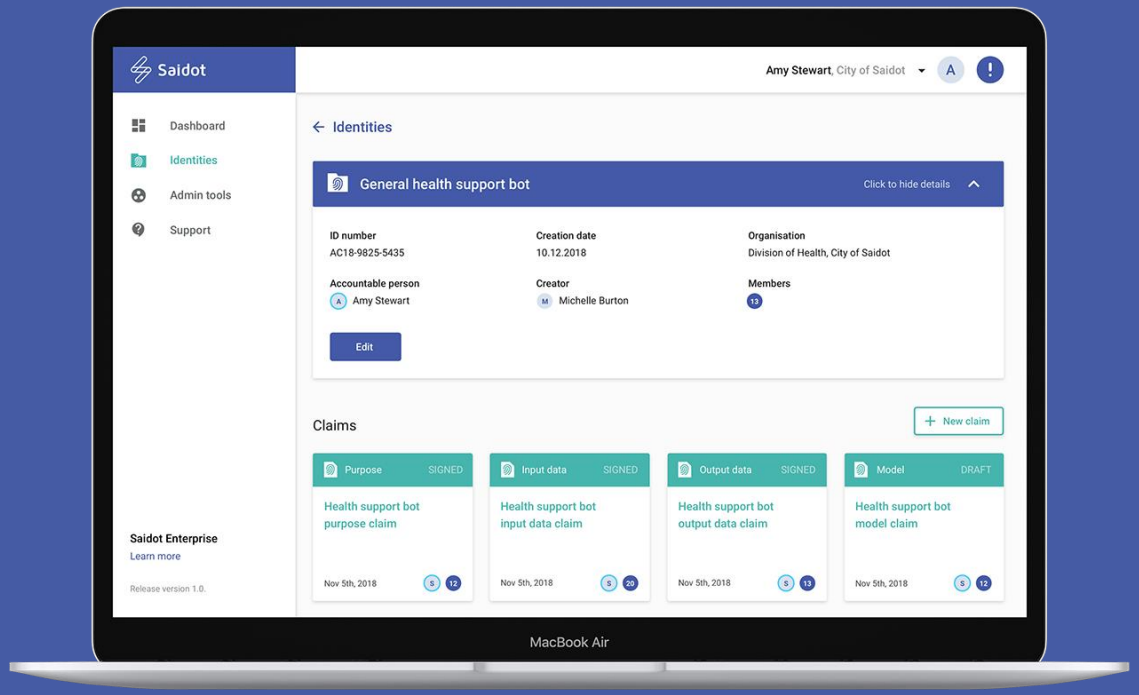
Need for algorithmic oversight

Fig. 2

From: Digital health: a path to validation



Saidot



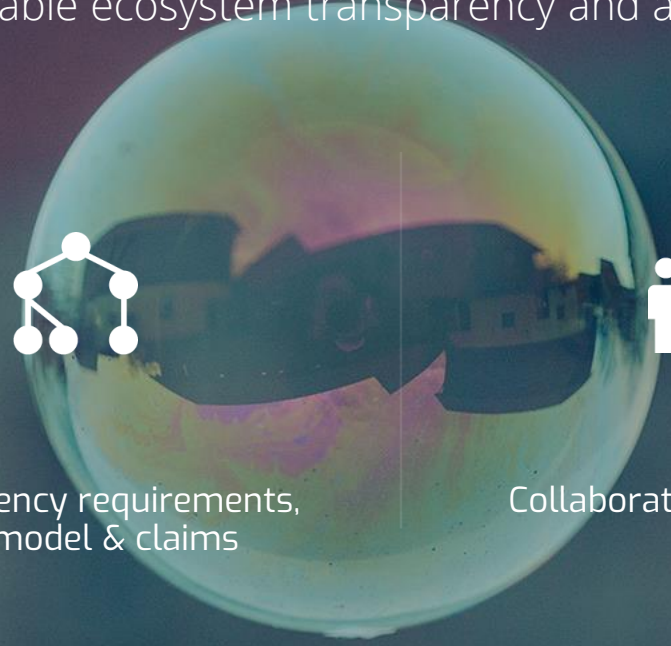
Technology for responsible AI ecosystems

Saidot enables ecosystem transparency and accountability on AI.

Saidot

Saidot AI transparency platform

We create technology to help companies & governments make, use, monetize and protect trustworthy & transparent AI. We enable ecosystem transparency and accountability of AI.



Registers for algorithmics systems

Transparency requirements, data model & claims

Collaboration with partners

3rd party interfaces

Saidot

Trust in public sector AI requires transparency and independent oversight.

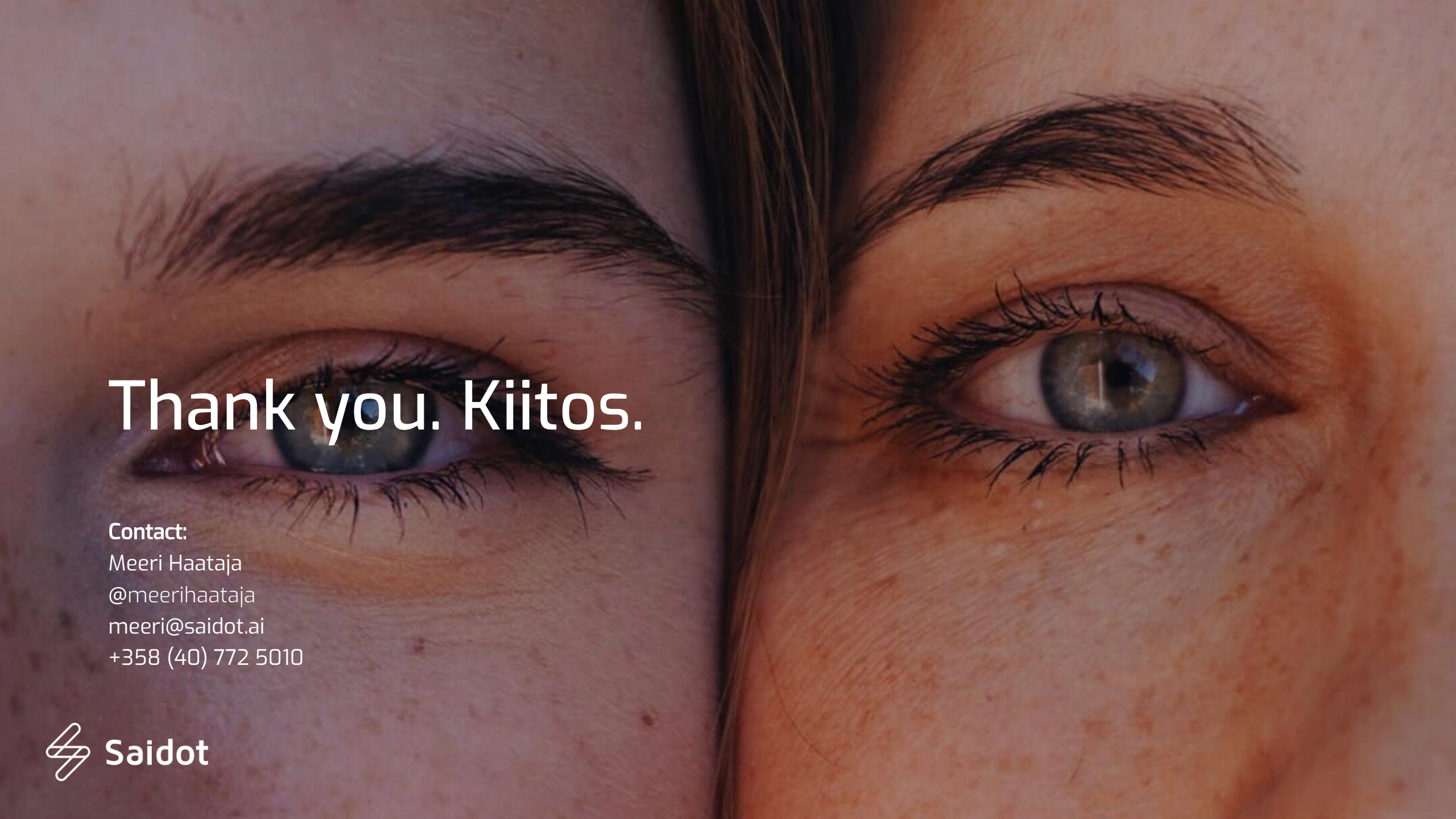
Saidot

Key takeaways

1. **Mistrust in public use of AI has led to extreme means of governing AI** especially in the facial recognition domain;
2. **Transparency is a foundational principle** that allows monitoring other ethics principles, and the ways public authority is exercised through algorithmic decisions;
3. Finland is leading the way for **creating global standards and benchmark** for public sector algorithmic openness & support for citizen agency;
4. Transparency is a means for oversight – we need to establish an **independent algorithmic oversight function** for the Finnish public sector.
5. Health professionals, **human operators of AI, are in a key role** demanding for algorithmic transparency

However

In future, the most unethical behavior may well be of not to consult AI for a second opinion in health care.



Thank you. Kiitos.

Contact:
Meeri Haataja
@meerihaataja
meeri@saidot.ai
+358 (40) 772 5010

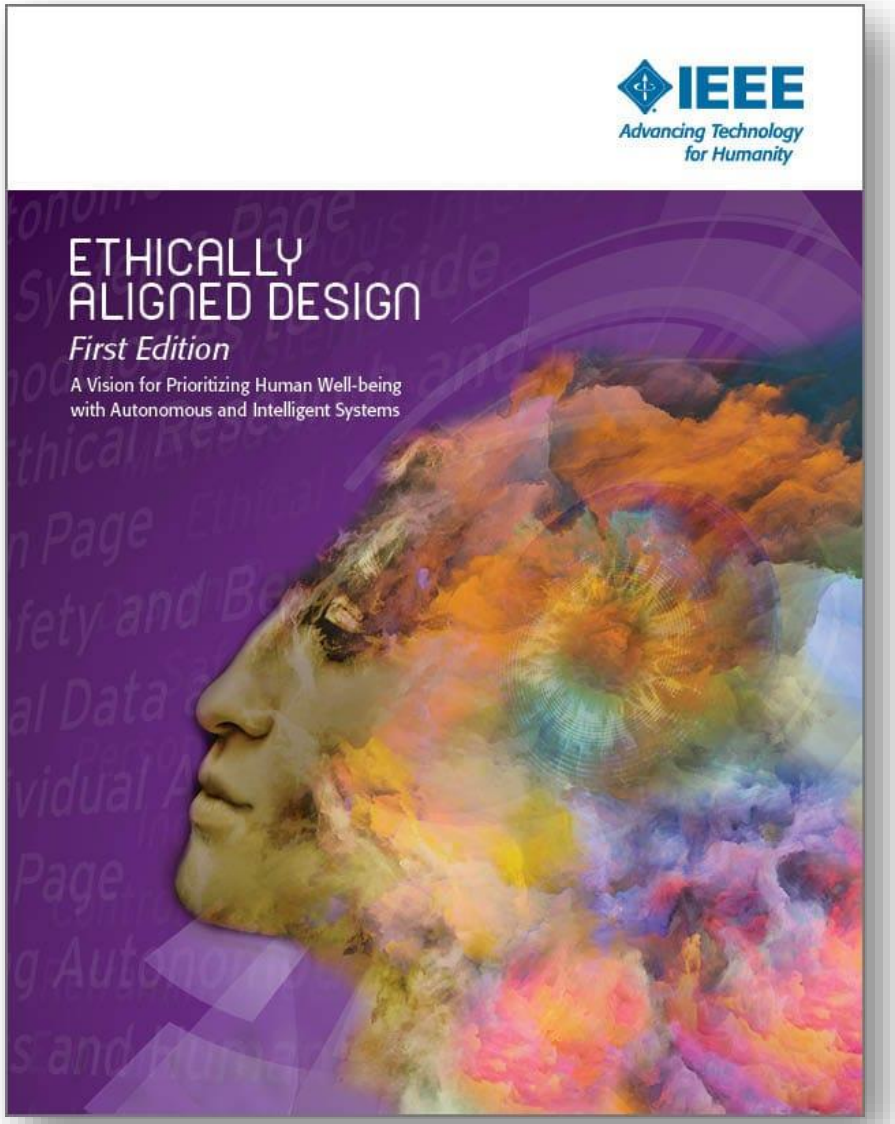
Saidot

Certifying AI

IEEE's ethics certifications for AI

The Ethics Certification Program for Autonomous and Intelligent Systems (ECPAIS) has the goal to create specifications for certification and marking processes that advance transparency, accountability, and reduction in algorithmic bias in autonomous and intelligent systems.

ECPAIS intends to offer a process and define a series of marks by which organizations can seek certifications for their processes around the A/IS products, systems, and services they provide.



Saidot

Certifying AI

IEEE's ethics certifications for AI

The Ethics Certification Program for Autonomous and Intelligent Systems (ECPAIS) has the goal to create specifications for certification and marking processes that advance transparency, accountability, and reduction in algorithmic bias in autonomous and intelligent systems.

ECPAIS intends to offer a process and define a series of marks by which organizations can seek certifications for their processes around the A/IS products, systems, and services they provide.

Types of information that should be considered in determining transparency demands in relation to a given A/IS		Stakeholders whose interest in access to different types of information should be considered in determining the transparency demands in relation to a given application of A/IS			
High-level category	**Specific type of information (examples) Disclosure of...**	**Operators**	**Decision-subjects**	**Public interest steward**	**General public**
Procedural aspects regarding A/IS employment and development	the fact that a given context involves the employment of A/IS	N/A	?	?	?
	how the employment of the system was authorized	?	?	?	?
	who developed the system	?	?	?	?
	...				
Data involved in A/IS development and operation	the origins of training data and data involved in the operation of the system	?	?	?	?
	the kinds of quality checks that data was subject to and their results	?	?	?	?
	how data labels are defined and to what extent data involves proxy variables	?	?	?	?
	relevant data sets themselves	?	?	?	?
	...				
Effectiveness/ performance	the kinds of effectiveness/performance measurement that have occurred	?	?	?	?
	measurement results	?	?	?	?
	any independent auditing or certification	?	?	?	?
	...				
Model specification	the input variables involved	?	?	?	?
	the variable(s) that the model optimizes for	?	?	?	?
	tthe complete model (complete formal representation, source code, etc.)	?	?	?	?
	...				
Explanation	information concerning the system's general logic or functioning	?	?	?	?
	information concerning the determinants of a particular output [125]	?	?	?	?
	...				

Saidot