Digital Trust Label & KI-Ethik-Label
Was bringen sie wirklich?

Niniane Päffgen & Lajla Fetic

#Shift2021

SHIFT 2021 – Labels und Deklarationen

Ziele der Breakout Session

Wir wollen:

- beleuchten, wieso die Entwicklung von Labels, Deklarationen und messbaren Regelwerken zu neuen Technologien wichtig und aktuell im Trend sind – und was wir uns davon erhoffen.
- hinterfragen, was Labels und Deklarationen in der Praxis bewirken können und wo die Grenzen sind.

Dafür:

- Zeigen wir mit zwei Beispielen aus der Praxis eine mögliche Umsetzung auf.
- Diskutieren wir mit den Teilnehmenden der SHIFT 2021 aus unterschiedlichen Perspektiven über die Umsetzbarkeit und Wirkung von Konsument:innen-Labels

LAJLA FETIC

Forscht und arbeitet zur Algorithmenethik, u.a. für die Bertelsmann Stiftung

Ist Co-Autorin mehrerer Praxisleitfäden zur Umsetzung von Ethik-Regeln

Mitglied der AI Ethics Impact Group

NINIANE PÄFFGEN

Leitet die Stiftung Swiss Digital Initiative mit Sitz in Genf

Fördert digitale Ethik und die Umsetzung ethischer Prinzipien in die Praxis und in der digitalen Welt

Arbeit am Digital Trust Label


CUT UP 5/4
F 4 3 2 0 ADD2
RATE SET
RATE CHECK
+TAX
−TAX
GT
ON
CA
MRC
M−
M+
C·CE
7
8
9
%
4
5
6
×
÷
1
2
3
+
−
0
00
=


NETFLIX

Digital Trust Label
Im Überblick

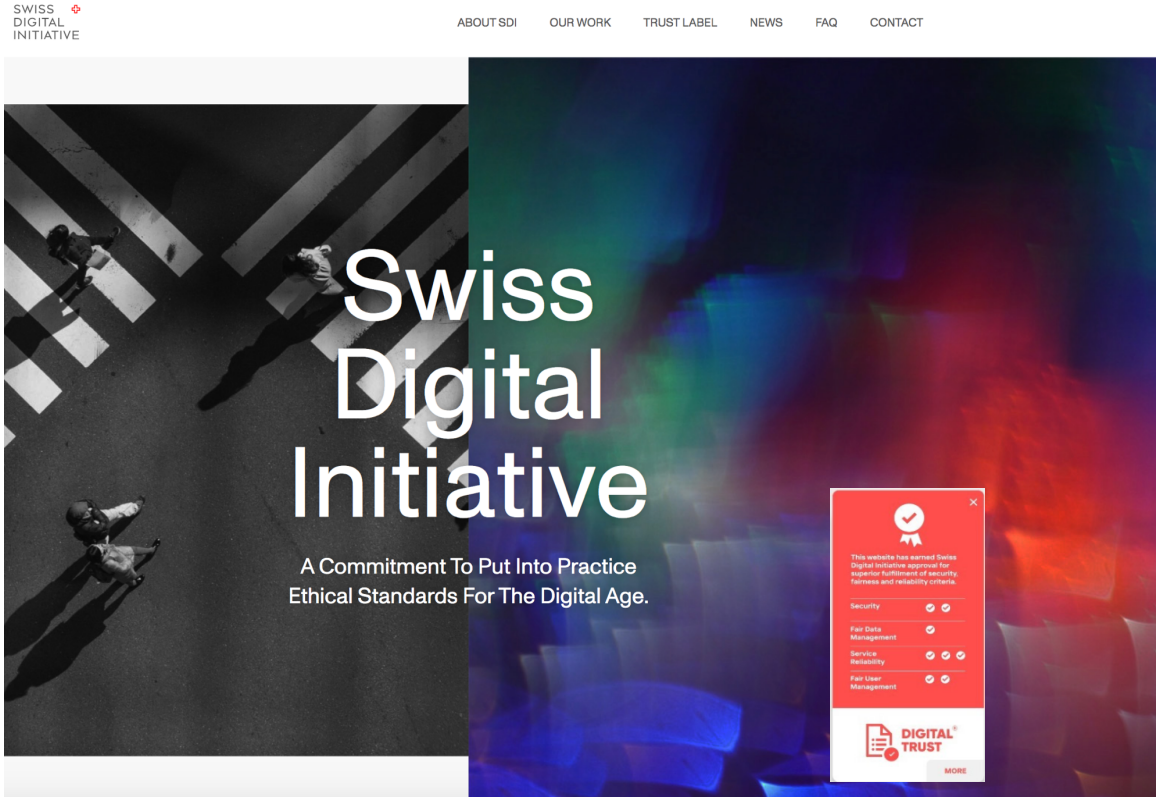
SWISS DIGITAL INITIATIVE

- Wir wollen das erste weltweit führende Digital Trust Label anbieten, welches die Vertrauenswürdigkeit eines digitalen Dienstes (Websites, Apps) in klarer, visuell verständlicher und nicht-technischer Sprache kennzeichnet.
- Das Label:
  - gibt Nutzerinnen und Nutzern digitaler Dienste **mehr Informationen und Transparenz**,
  - zeigt, dass ein Unternehmen **die Verantwortung gegenüber den Nutzerinnen und Nutzern** seiner digitalen Dienste ernst nimmt,
  - umfasst 4 Kategorien: **Sicherheit, Fair Data Management, Verlässlichkeit des Services, Fair User Management,**
  - besteht aus technischen Kriterien, die von einer **unabhängigen dritten Partei verifiziert und auditiert werden**.

BEISPIELILLUSTRATION

Digital Trust Label

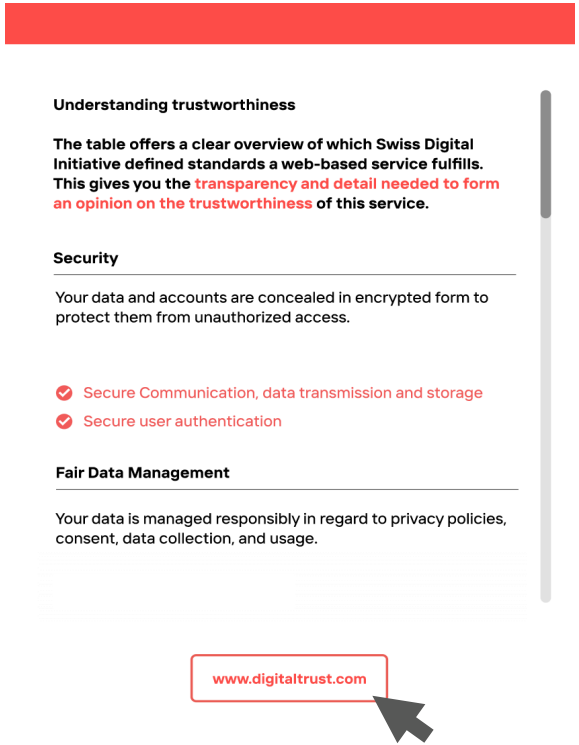
Beispiel User Journey (noch nicht final) – Ebene I



Digital Trust Label

Beispiel User Journey (noch nicht final) – Ebene II

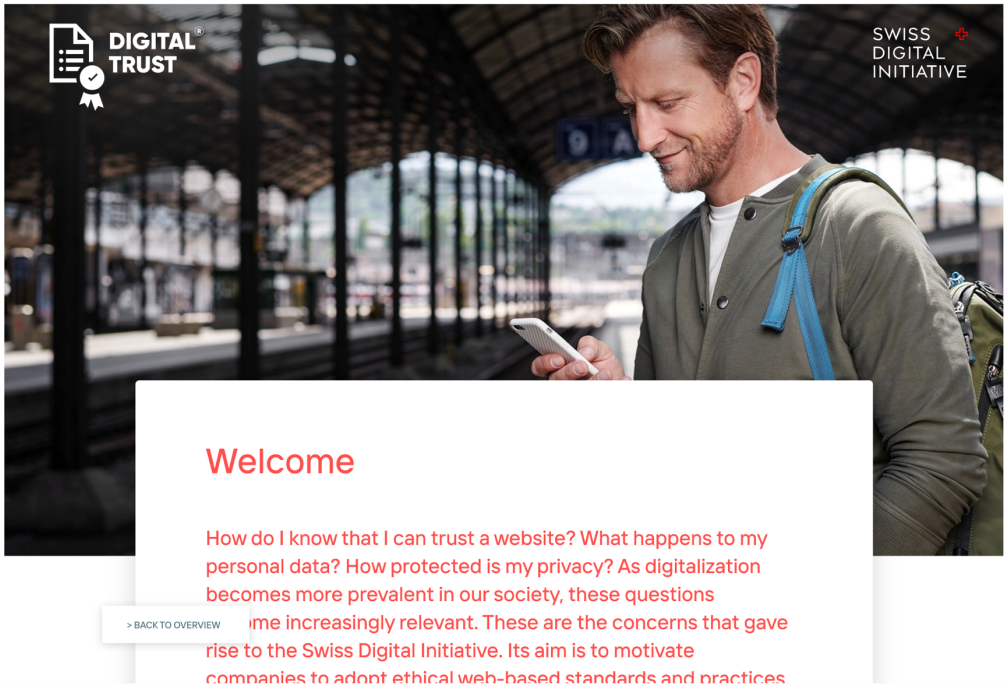
SWISS DIGITAL INITIATIVE



Digital Trust Label

Beispiel User Journey (noch nicht final) – Ebene III

SWISS DIGITAL INITIATIVE



Digital Trust Label

Warum ein Label – Was erhoffen wir uns davon?

SWISS DIGITAL INITIATIVE

- **Mehr Informationen und Transparenz** für Nutzerinnen und Nutzer digitaler Dienste.
- Nutzerinnen und Nutzer können die **Vertrauenswürdigkeit digitaler Dienste besser einschätzen**.
- Ein Label **reduziert technische Komplexität und sensibilisiert in einer nutzerfreundlichen Sprache**: Hat die Nutzerin beispielsweise mit einem Algorithmus zu tun?
- Es steht für einen **Benchmark und signalisiert, dass Unternehmen die Vertrauenswürdigkeit ihrer digitalen Dienste** ernst nehmen.
- Die Service Provider nehmen durch die Anwendung der Labelkriterien **ihre Verantwortung gegenüber den Nutzerinnen und Nutzern digitaler Services** wahr.
- **Werte messbar** und von der abstrakten Theorie in die Praxis überführen.

Digital Trust Label - Erfolgskriterien

Erkenntnisse aus dem bisherigen Prozess

SWISS DIGITAL INITIATIVE

1. **Globale Nutzerstudie:** Hohe (globale) Nachfrage nach einer Lösung zu Vertrauen im Internet besteht.
2. **Inklusiver und transparenter Multi-Stakeholder** Entwicklungs-Prozess sind zentral
3. Das Label wird auch nach dem Launch **ständig weiterentwickelt** werden müssen
4. Das Label muss eine Kombination aus **normativen** (Garantie eines gewissen Standards) und **deskriptiven Elementen** (Informationen) sein.
5. Digitales Vertrauen kann **nicht rein technisch** gelöst werden: Zusammenspiel verschiedener Faktoren (Kontext, offline Erfahrungen etc.)
6. Die **Einhaltung und Erfüllung der Kriterien** müssen durch eine unabhängige Organisation **geprüft** werden.

Digital Trust Label

50 + Initiativen weltweit mit ähnlichem Ziel

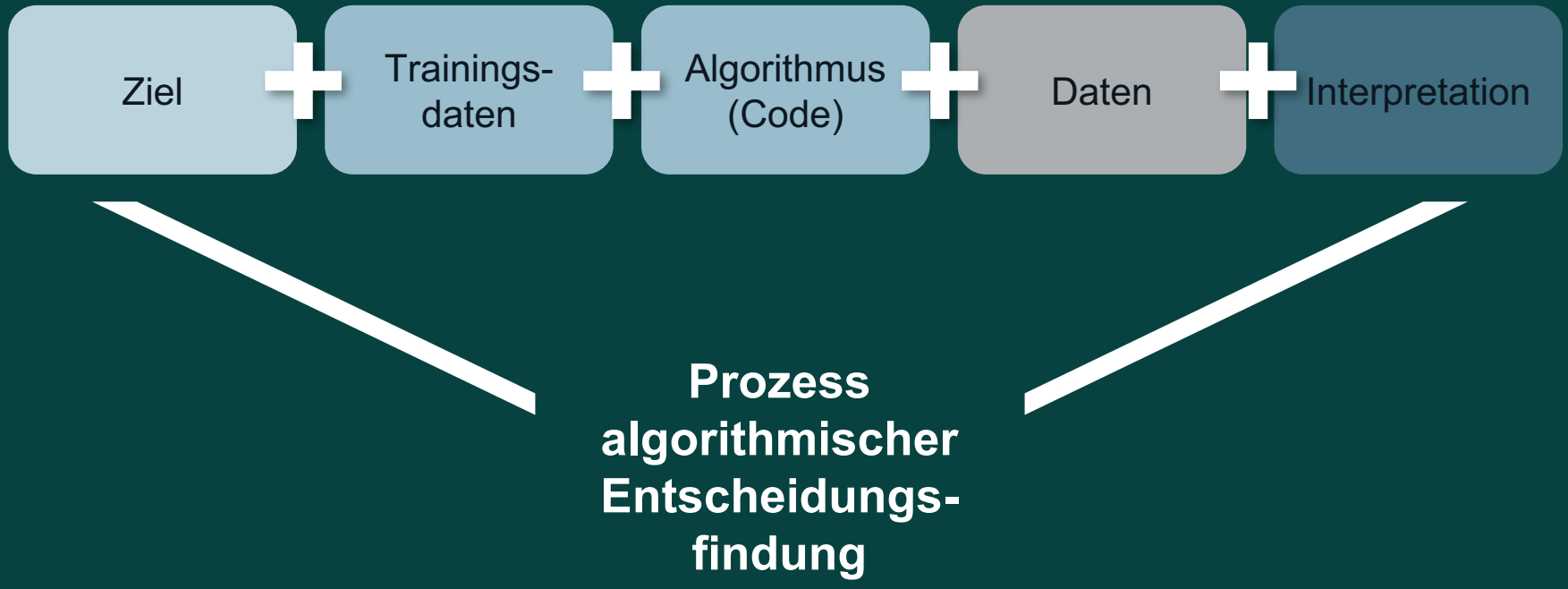
SWISS DIGITAL INITIATIVE

Trustmark for the Internet (by NGI.eu)	Regional	Concept (stage 1)	Label (end user)	Next Generation Internet (NGI)	EU
A Trustworthy Tech Mark	International	Abandoned	Label (end user)	DotEveryone	Association
AI Certification	National	Prototype (stage 3)	Certification	Fraunhofer Institute, Germany's Federal Office f...	Academia Governement
D-Seal, Seal for Data Ethics	National	Running (stage 4)	Label (end user)	Danish Minister for Industry, Business and Financial ...	Governement
Data Ethics Framework	Regional	Realisation (stage 2)	Label (end user)	Bertelsmann Stiftung + AI Ethics Impact Group	Foundation
IHAN (Human-Driven Data Economy)	Regional	Realisation (stage 2)	Ecosystem	Sitra	Foundation Governement
Independent Audit of AI Systems	International	Prototype (stage 3)	Label (end user)	For Humanity	Foundation
The Digital Standard	International	Running (stage 4)	Evaluation	Consumer Reports, Disconnect, Ranking Digital...	Association
The Ethics Certification Program for Autonomous ...	International	Prototype (stage 3)	Label (end user)	IEEE	International Organizati...

„KI ist sehr gut darin, die Welt zu beschreiben, so wie sie heute ist, mit all ihren Vorurteilen. Aber KI weiß nicht, wie die Welt sein sollte.“

Joanne Chen, Partner bei Foundation Capital

Algorithmische Entscheidungsprozesse sind von Mensch-Technik-Interaktionen geprägt.



Die 9 Algo.Rules – Den Prozess im Blick

#1 Kompetenz aufbauen

#2 Verantwortung definieren

#3 Ziele und erwartete Wirkung dokumentieren

#4 Sicherheit gewährleisten

#5 Kennzeichnung durchführen

#6 Nachvollziehbarkeit sicherstellen

#7 Beherrschbarkeit absichern

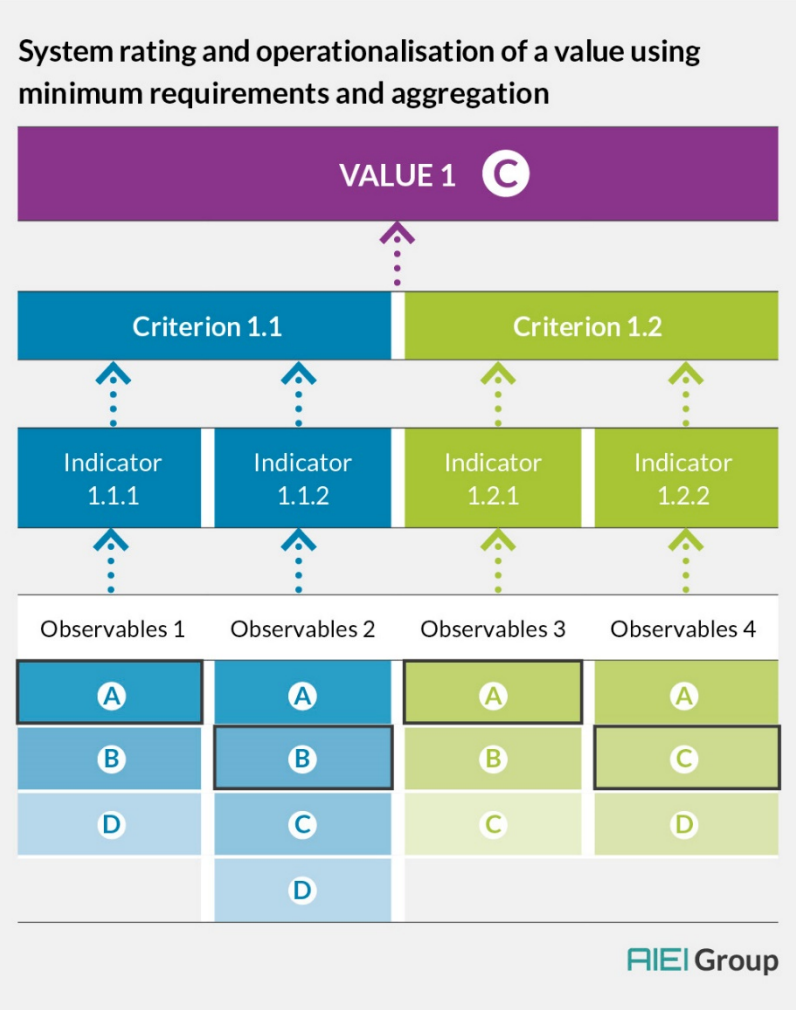
#8 Wirkung überprüfen

#9 Beschwerden ermöglichen

Nach einer Studie der ETH Zürich (Jobin et. al. 2019) finden sich bestimmte ethische Prinzipien immer wieder in den Ethik-Richtlinien.

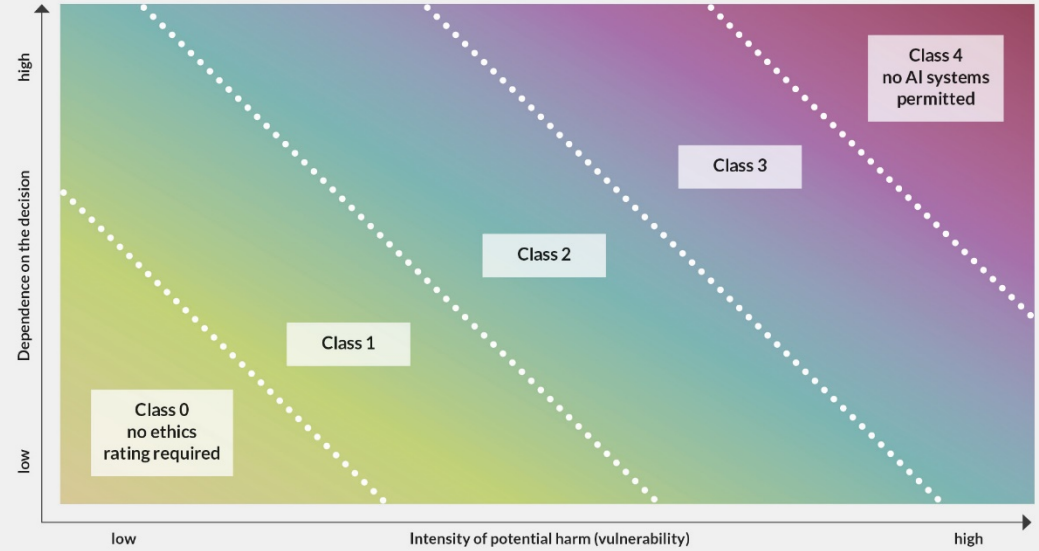
SPANNEND:
Die Definitionen der Prinzipien fehlen, unterscheiden oder widersprechen sich teilweise.

Fairness
Accountability
Transparenz:
Was bedeutet das überhaupt?



Der Weg von abstrakten Werten zu messbaren Indikatoren ist kompliziert.

Risk matrix with 5 classes of application areas with risk potential ranging from 'no ethics rating required' in class 0 to the prohibition of AI systems in class 4

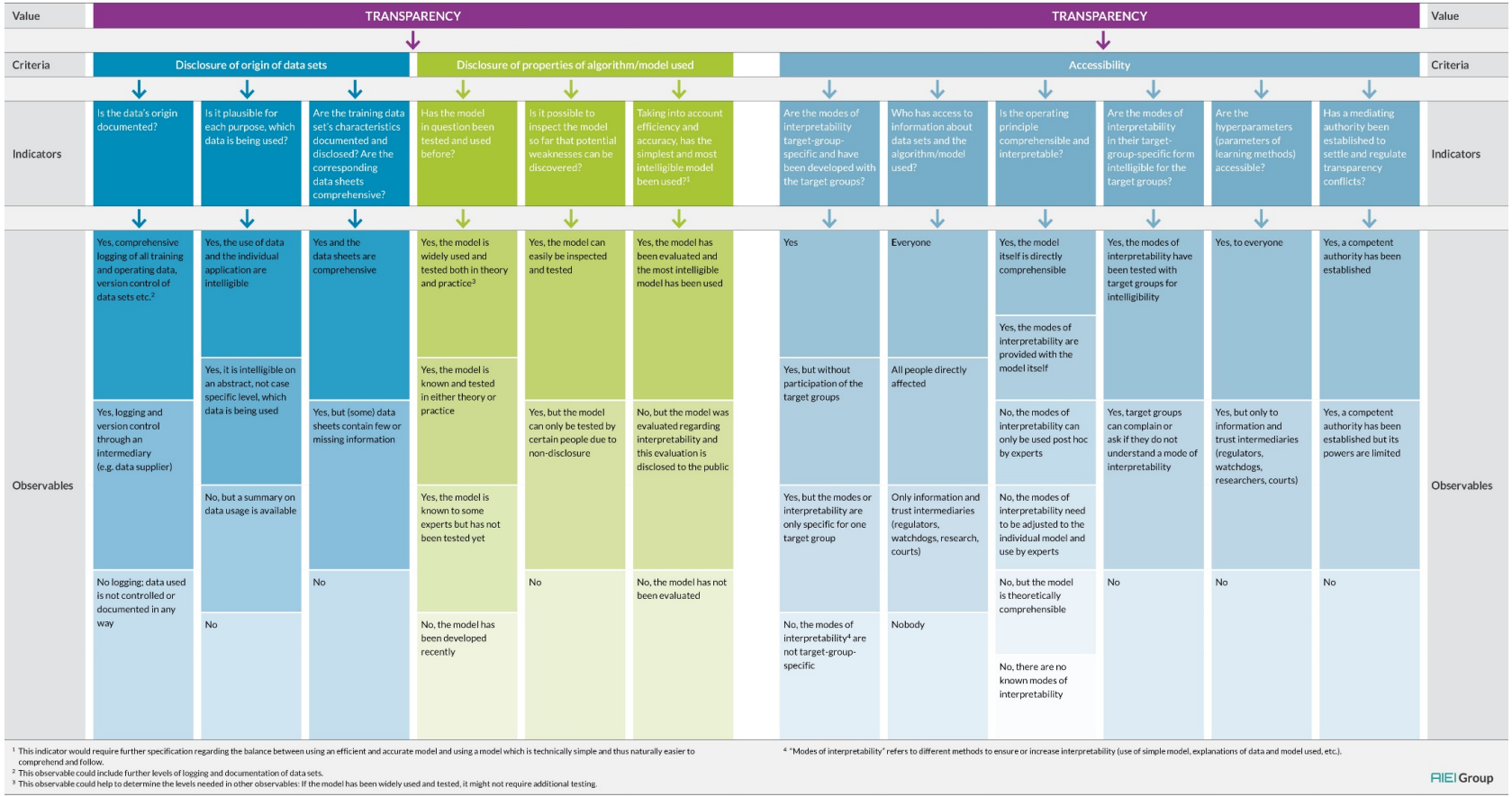


Source: Krafft and Zweig 2019

AIEI Group

Die Anforderungen sind nicht für jeden Anwendungs-kontext gleich.

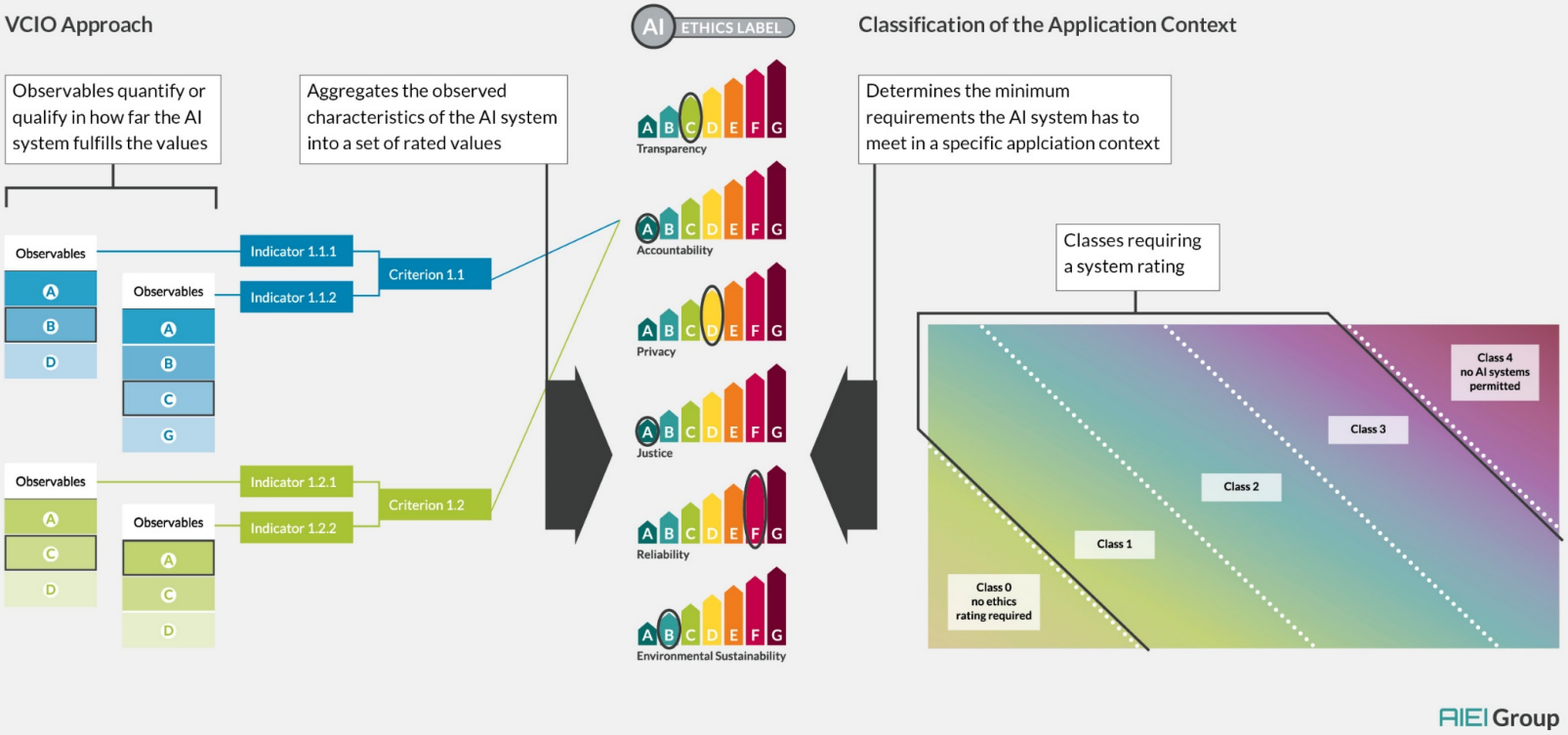
Applying the VCIO approach to transparency as a value



[1] This indicator would require further specification regarding the balance between using an efficient and accurate model and using a model which is technically simple and thus naturally easier to comprehend and follow.
[2] This observable could include further levels of logging and documentation of data sets.
[3] This observable could help to determine the levels needed in other observables: If the model has been widely used and tested, it might not require additional testing.
[4] "Modes of interpretability" refers to different methods to ensure or increase interpretability (use of simple model, explanations of data and model used, etc.).

AIEI Group

Illustration of the composition of the whole system rating using minimum requirements



Reicht das KI-Ethik-Label aus?

Wir müssen die Delegation von Verantwortung verhindern.

Label ersetzen die gesellschaftliche Debatte über den Einsatz und die Anwendung von Technologie nicht.

Wir brauchen einen Werkzeugkasten für KI-Ethik und eine systemische Perspektive.

SWISS DIGITAL INITIATIVE

Niniane Päffgen
Geschäftsführerin
Swiss Digital Initiative

niniane@sdi-foundation.org

c/o Campus Biotech, Chemin des Mines 9, 1202 Genève

https://swiss-digital-initiative.org

L
F

Lajla Fetic
Wissenschaftlerin und Beraterin
KI-Ethik und Regulierung

lajla.fetic@email.de

@lajlafetic
www.linkedin.com/in/lajlafetic/
www.lajlafetic.de