



Normative Implications of Ethical Guidelines

1 Nov, 2022

[Stefan Larsson](#)

Associate Professor in
Technology and Social Change,
Department of Technology and
Society, LTH, Lund University

Lawyer (LLM)
PhD in sociology of law
PhD in spatial planning fysisk planering



LUNDS
UNIVERSITET

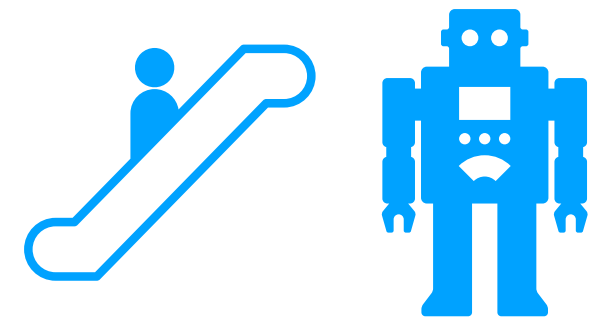
Focus for Research Group



Stefan Larsson

Associate Professor in
Technology and Social Change,
Department of Technology and
Society, LTH, Lund University

Lawyer (LLM)
PhD in sociology of law
PhD in spatial planning fysisk planering

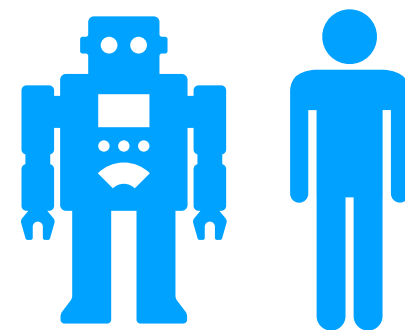


AI Transparency and
Consumer Trust



Transparency and
medical AI:

- AIR Lund (registry based)
- MASAI (mammography)



Socio-legal
robotics



ADM-GOV: Governance
of AI & ADM in the
public sector

- Swe-Fin



The Automated Administration:
Governance of ADM in the public sector



WP 1: Principled AI

Stefan & Jockum

How is the European principled approach to AI governance understood in the Nordic public administration and what are the implications for ADM deployment?

FROM ETHICAL
PRINCIPLES TO
NORMATIVE
IMAGINARIES

In short

1. Ethics & AI: [A Formative Period](#)
2. Ethics & AI: [European Member States](#)
3. Ethics & AI: [Sweden](#)
4. Ongoing Study: [AI and Discrimination](#)

Drawing from

- Larsson, S. & Ledendal, J. (2022) "AI i offentlig sektor: Från etiska riktlinjer till lagstiftning", in de Vries & Dahlberg (eds.) **Law, AI & Digitalization. De Lege – Yearbook Uppsala Faculty of Law 2021**. Uppsala: Iustus Förlag.
- Larsson, S. (2020) On the Governance of Artificial Intelligence through Ethics Guidelines, *Asian Journal of Law and Society*, 7(1): 1-23.
- Larsson, S. & Heintz, F. (2020) Transparency in artificial intelligence, *Internet Policy Review* 9(2): 1-16.
- Larsson, S., Ingram Bogusz, C., & Andersson Schwarz, J. Eds. (2020) **Human-Centred AI in the EU. Trustworthiness as a strategic priority in the European Member States**. Brussels: European Liberal Forum.

*On the Governance of Artificial Intelligence through Ethics Guidelines**

Stefan Larsson & Jonas Ledendal

AI i offentlig sektor: Från etiska riktlinjer till lagstiftning

1 Inledning

Användningen av artificiell intelligens (AI), det vill säga en rad primärt databeroende metoder och teknologier för bland annat prediktion och automation, tycks bidra till en omvälvande period i den offentliga förvaltningen. Detta skifte föranleder ett styrningsbehov, ofta kopplat till tillitsfrågor, vilket i en internationell kontext inte minst utgör en mängd etiska riktlinjer och principiella dokument som publicerade senaste åren. Styrningen befinner sig därmed i en form vilket inte minst tydliggörs i och med EU-kommissionens förordning om harmoniserade regler för AI (rättsakt om AI) som antogs i april 2021.¹ Vi vill här belysa denna styrningsutveckling och visa på dess mest centrala delar, samt analysera den svenska AI-användningen i dess ljus.

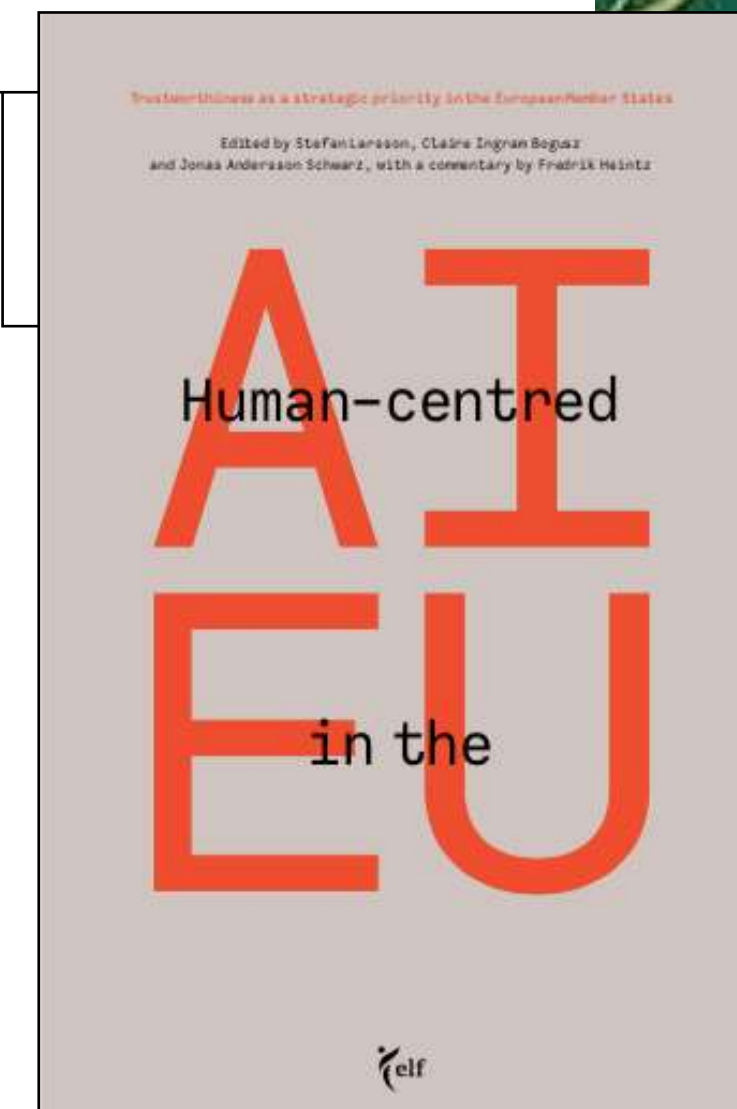
1.1 Regeringsuppdrag i ljuset av internationell AI

I juni 2021 fick Myndigheten för digital förvaltning (DIGG), Myndigheten för samhällsplanering och byggnadsverk, och Skatteverket i uppdrag av regeringen att främja offentlig förvaltnings förmåga att använda artificiell intelligens (AI) i syfte att stärka Sveriges välfärd och konkurrenskraft.

¹ Europeiska kommissionen, Förslag till Europaparlamentets och rådets förordning om harmoniserade regler för artificiell intelligens (rättsakt om artificiell intelligens) (2021/01061), 21.4.2021, COM(2021) 206 final.

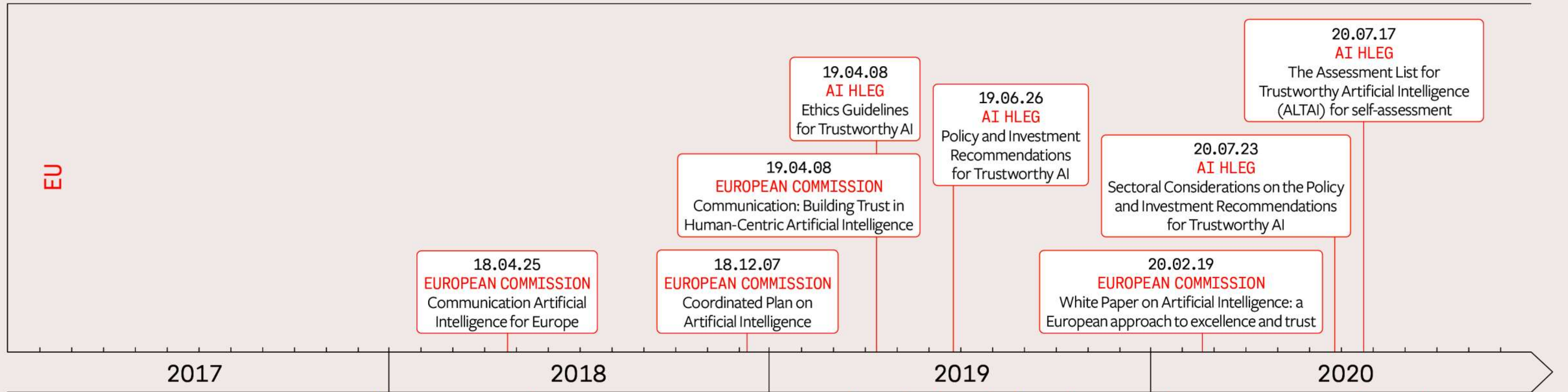
² Regeringen (21 juni 2021) "Uppdrag att främja offentlig förvaltnings förmåga att använda artificiell intelligens". Diarienummer: I2021/01825.

of ethics guidelines as a governance tool. This has become a central policy instrument, as well as the EU, focused on the Ethics Guidelines for Trustworthy Artificial Intelligence published by the EU Commission in 2020. This article addresses the gap between technology and law by overlapping and already-existing instruments such as the article concerning the governance of AI and (3)



Ethics & AI: A Formative Period

AI in the EU



Guidelines: Global Outlook

DOCUMENT TIMELINE



2017

2018

2019

2020



PRINCIPLED ARTIFICIAL INTELLIGENCE

A Map of Ethical and Rights-Based Approaches to Principles for AI

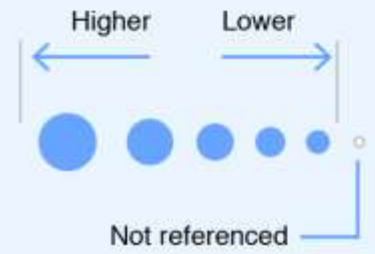
Authors: Jessica Fjeld, Nele Achten, Hannah Hilligoss, Adam Nagy, Madhulika Srikumar

Designers: Arushi Singh (arushisingh.net) and Melissa Axelrod (melissaaxelrod.com)

HOW TO READ:

Date, Location
Document Title
Actor

COVERAGE OF THEMES:



The size of each dot represents the percentage of principles in that theme contained in the document. Since the number of principles per theme varies, it's informative to compare dot sizes within a theme but not between themes.

The principles within each theme are:

Privacy:

- Privacy
- Control over Use of Data
- Consent
- Privacy by Design
- Recommendation for Data Protection Laws
- Ability to Restrict Processing
- Right to Rectification
- Right to Erasure

Accountability:

- Accountability
- Recommendation for New Regulations
- Impact Assessment
- Evaluation and Auditing Requirement
- Verifiability and Replicability
- Liability and Legal Responsibility
- Ability to Appeal
- Environmental Responsibility
- Creation of a Monitoring Body
- Remedy for Automated Decision

Safety and Security:

- Security
- Safety and Reliability
- Predictability
- Security by Design

Transparency and Explainability:

- Explainability
- Transparency
- Open Source Data and Algorithms
- Notification when Interacting with an AI
- Notification when AI Makes a Decision about an Individual
- Regular Reporting Requirement
- Right to Information
- Open Procurement (for Government)

Fairness and Non-discrimination:

- Non-discrimination and the Prevention of Bias
- Fairness
- Inclusiveness in Design
- Inclusiveness in Impact
- Representative and High Quality Data
- Equality

Human Control of Technology:

- Human Control of Technology
- Human Review of Automated Decision
- Ability to Opt out of Automated Decision

Professional Responsibility:

- Multistakeholder Collaboration
- Responsible Design
- Consideration of Long Term Effects
- Accuracy
- Scientific Integrity

Promotion of Human Values:

- Leveraged to Benefit Society
- Human Values and Human Flourishing
- Access to Technology

Further information on findings and methodology is available in *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches* (Berkman Klein, 2020) available at cyber.harvard.edu.



Most common principles

- 2016–2019: At least 84 initiatives with principled or ethical guidelines for AI
- Much convergence around i) transparency, ii) justice / fairness, iii) non-maleficence, iv) responsibility v) privacy
- ”Substantive divergence” in relation to how these principles are interpreted

nature machine intelligence PERSPECTIVE
<https://doi.org/10.1038/s42256-019-0088-2>

The global landscape of AI ethics guidelines

Anna Jobin, Marcello Lenca and Effy Vayena*

In the past five years, private companies, research institutions and public sector organizations have issued principles and guidelines for ethical artificial intelligence (AI). However, despite an apparent agreement that AI should be ‘ethical’, there is debate about both what constitutes ‘ethical AI’ and which ethical requirements, technical standards and best practices are needed for its realization. To investigate whether a global agreement on these questions is emerging, we mapped and analysed the current corpus of principles and guidelines on ethical AI. Our results reveal a global convergence emerging around five ethical principles (transparency, justice and fairness, non-maleficence, responsibility and privacy), with substantive divergence in relation to how these principles are interpreted, why they are deemed important, what issue, domain or actors they pertain to, and how they should be implemented. Our findings highlight the importance of integrating guideline-development efforts with substantive ethical analysis and adequate implementation strategies.

Artificial intelligence (AI), or the theory and development of computer systems able to perform tasks normally requiring human intelligence, is widely heralded as an ongoing “revolution” transforming science and society altogether^{1,2}. While approaches to AI such as machine learning, deep learning and artificial neural networks are reshaping data processing and analysis³, autonomous and semi-autonomous systems are being increasingly used in a variety of sectors including healthcare, transportation and the production chain⁴. In light of its powerful transformative force and profound impact across various societal domains, AI has sparked ample debate about the principles and values that should guide its development and use⁵. Fears that AI might jeopardize jobs for human workers⁶, be misused by malevolent actors⁷, elude accountability or inadvertently disseminate bias and thereby undermine fairness⁸ have been at the forefront of the recent scientific literature and media coverage. Several studies have discussed the topic of ethical AI^{9–12}, notably in meta-assessments^{13–15} or in relation to systemic risks^{16,17} and unintended negative consequences such as algorithmic bias or discrimination^{18–21}.

National and international organizations have responded to these concerns by developing ad hoc expert committees on AI, often mandated to draft policy documents. These committees include the High-Level Expert Group on Artificial Intelligence appointed by the European Commission, the expert group on AI in Society of the Organisation for Economic Co-operation and Development (OECD), the Advisory Council on the Ethical Use of Artificial Intelligence and Data in Singapore, and the Select Committee on Artificial Intelligence of the UK House of Lords. As part of their institutional appointments, these committees have produced or are reportedly producing reports and guidance documents on AI. Similar efforts are taking place in the private sector, especially among corporations who rely on AI for their business. In 2018 alone, companies such as Google and SAP publicly released AI guidelines and principles. Declarations and recommendations have also been issued by professional associations and non-profit organizations such as the Association of Computing Machinery (ACM), Access Now and Amnesty International. This proliferation of soft-law efforts can be interpreted as a governance response to advanced research into AI, whose research output and market size have drastically increased²² in recent years.

Reports and guidance documents for ethical AI are instances of what is termed non-legislative policy instruments or soft law²³. Unlike so-called hard law—that is, legally binding regulations passed by the legislatures to define permitted or prohibited conduct—ethics guidelines are not legally binding but persuasive in nature. Such documents are aimed at assisting with—and have been observed to have significant practical influence on—decision-making in certain fields, comparable to that of legislative norms²⁴. Indeed, the intense efforts of such a diverse set of stakeholders in issuing AI principles and policies is noteworthy, because they demonstrate not only the need for ethical guidance, but also the strong interest of these stakeholders to shape the ethics of AI in ways that meet their respective priorities^{25–27}. Specifically, the private sector’s involvement in the AI ethics arena has been called into question for potentially using such high-level soft policy as a portmanteau to either render a social problem technical²⁸ or to eschew regulation altogether²⁹. Beyond the composition of the groups that have produced ethical guidance on AI, the content of this guidance itself is of interest. Are these various groups converging on what ethical AI should be, and the ethical principles that will determine the development of AI? If they diverge, what are their differences and can these differences be reconciled?

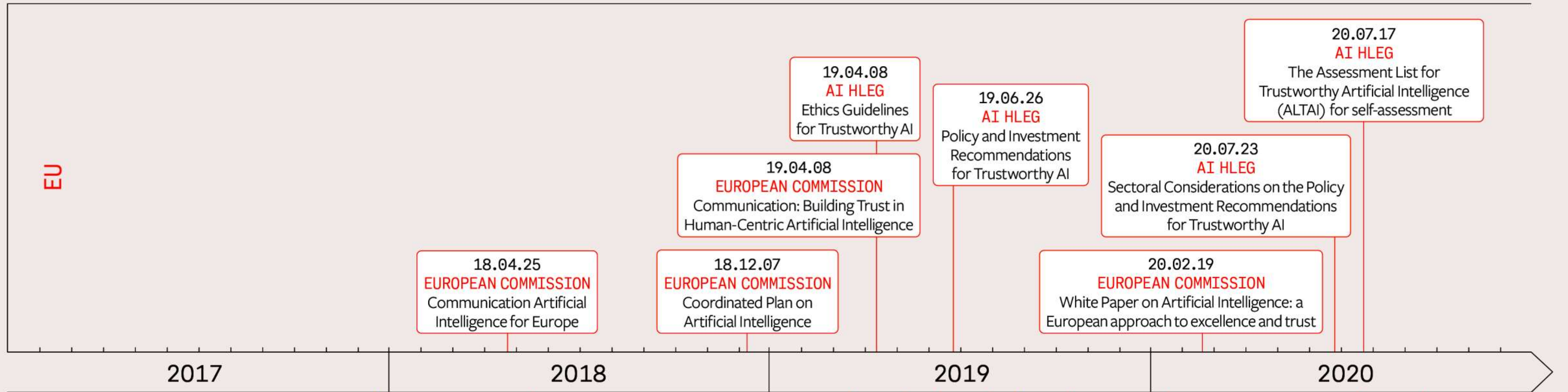
Our Perspective maps the global landscape of existing ethics guidelines for AI and analyses whether a global convergence is emerging regarding both the principles for ethical AI and the suggestions regarding its realization. This analysis will inform scientists, research institutions, funding agencies, governmental and intergovernmental organizations, and other relevant stakeholders involved in the advancement of ethically responsible innovation in AI.

Methods
We conducted a scoping review of the existing corpus of documents containing soft-law or non-legal norms issued by organizations. This included a search for grey literature containing principles and guidelines for ethical AI, with academic and legal sources excluded. A scoping review is a method aimed at synthesizing and mapping the existing literature³⁰ that is considered particularly suitable for complex or heterogeneous areas of research^{31,32}. Given the absence of a unified database for AI-specific ethics guidelines, we developed a protocol for discovery and eligibility, adapted from the Preferred

Health Ethics and Policy Lab, Department of Health Sciences and Technology, ETH Zurich, Zurich, Switzerland. *e-mail: effy.vayena@hest.ethz.ch

NATURE MACHINE INTELLIGENCE | VOL 1 | SEPTEMBER 2019 | 389–399 | www.nature.com/naturemachintell 389

AI in the EU



INDEPENDENT
HIGH-LEVEL EXPERT GROUP ON
ARTIFICIAL INTELLIGENCE
SET UP BY THE EUROPEAN COMMISSION



THE ASSESSMENT LIST FOR
TRUSTWORTHY ARTIFICIAL
INTELLIGENCE (ALTAI)
for self assessment

REQUIREMENT #1 Human Agency and Oversight

Human Agency and Autonomy
Human Oversight

REQUIREMENT #2 Technical Robustness and Safety

Resilience to Attack and Security
General Safety
Accuracy
Reliability, Fall-back plans and Reproducibility

REQUIREMENT #3 Privacy and Data Governance

Privacy
Data Governance

REQUIREMENT #4 Transparency

Traceability
Explainability
Communication

REQUIREMENT #5 Diversity, Non-discrimination and Fairness

Avoidance of Unfair Bias
Accessibility and Universal Design
Stakeholder Participation

REQUIREMENT #6 Societal and Environmental Well-being

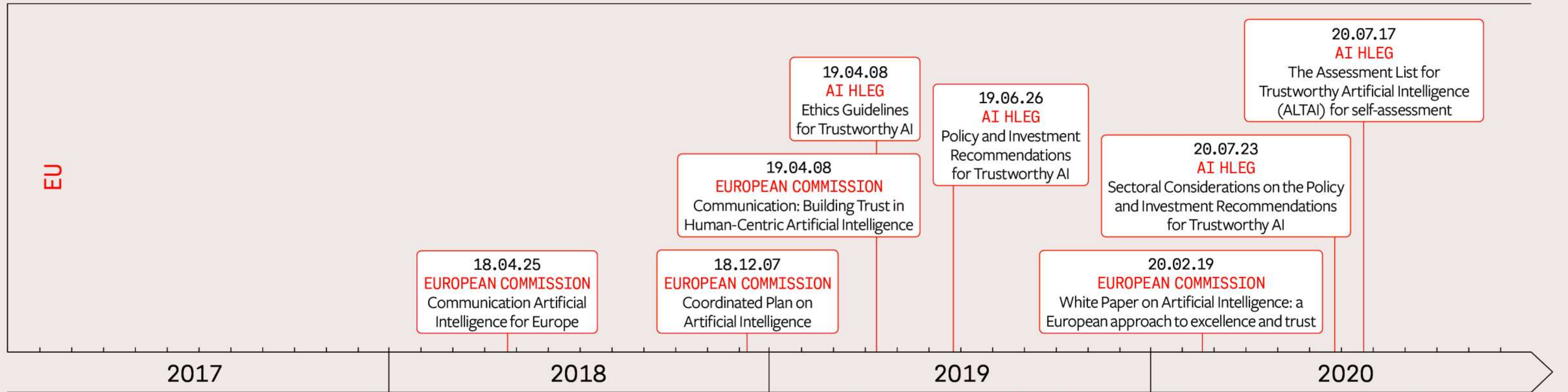
Environmental Well-being
Impact on Work and Skills
Impact on Society at large or Democracy

REQUIREMENT #7 Accountability

Auditability
Risk Management

Ethics & AI: European Member States

AI in the EU



AI policy in Portugal

Ambitious, yet laconic about legal routes
towards trustworthy AI

AI PORTUGAL 2030 (June 2019), “ethical-by-design”: privacy, safety, transparency, fairness and inclusion. Education, civic empowerment.
“thin” on law.

AI policy in Norway

Looking to the future and harmonised with the
EU

Jan 2020, clear influence from Ethics
guidelines.

AI policy in Poland

Ethical considerations already at the core
3 docs, last draft in Sep 2020; Conflicting
perspectives in politics and AI strategies;
Includes: funding, innovation, education,
consumers, ethical principles

AI policy in the Nordics

Pledging openness, transparency and trust, while
expressing readiness to apply AI in society
cultural unity; themes of democracy, ethics and
privacy are also prominent

AI policy in the Netherlands

More focus on practice than principles when it comes to trustworthiness

Eager on AI; critique on shaky public funding; strategy not clearly reflecting HLEG's; a bit scattered documents. 2 interesting cases.

AI policy in the Czech Republic

Strong business focus, welcoming towards foreign investment

May 2019, not explicitly linked to principles in Ethics guidelines; focus economic growth, industry.

AI policy in Italy

Comprehensive focus on core infrastructural robustness and humanistic values

July 2020, reflecting Ethics Guidelines, digitalisation of public administration,

Ethics & AI: Sweden

- **DIGG 2020:** can save “140 billion SEK annually”, we should
 - Establish a competence center with expertise in AI;
 - Develop a platform for collaboration, co-development and innovation
 - Produce an AI guide
 - Create legal conditions to facilitate experimental activities
 - Develop a national data strategy for public administration
- **DIGG & Land Survey Authorities, 2020:** Trust model, documentation, open log; AI registry need to be investigated.



Government mandate: Promote AI in public admin!

- Four big authorities: **Develop an “AI guide”** for public administration
- ... to be “adapted to relevant international recommendations and guidelines for the AI field” and
- Develop a trust model for automated decision making supported by AI
- To be reported Jan 2023.



The image shows a document from the Swedish Government (Regeringen) regarding an AI mandate. The document is titled "Regeringsbeslut" and is dated 2021-06-17 with the reference number I2021/01825. It is issued by the Infrastructure Department (Infrastrukturdepartementet). The document contains a mandate to develop an AI guide for public administration, adapted to international recommendations and guidelines. The mandate is to be reported by January 2023. The document also includes a section titled "Regeringen beslut" which details the government's decision to task the Swedish Public Employment Agency (Arbetsförmedlingen), the Swedish Companies Registration Office (Bolagsverket), the Swedish Agency for Digital Government (Myndigheten för digital förvaltning (Digg)), and the Swedish Tax Authority (Skatteverket) to develop an AI guide for public administration, with the goal of strengthening Sweden's welfare and competitiveness. The document also mentions that Digg should coordinate the work of the agencies.

Regeringen

Regeringsbeslut II 6
1 bilaga

2021-06-17
I2021/01825

Infrastrukturdepartementet

Adressater: Se bilaga

Uppdrag att främja offentlig förvaltnings förmåga att använda artificiell intelligens

Regeringen beslut

Regeringen uppdrar åt Arbetsförmedlingen, Bolagsverket, Myndigheten för digital förvaltning (Digg) och Skatteverket (myndigheterna) att enligt vad som anges under rubriken Närmare om uppdraget främja offentlig förvaltnings förmåga att använda artificiell intelligens (AI) i syfte att stärka Sveriges välfärd och konkurrenskraft. Digg ska samordna myndigheternas arbete.

WP 1: Principled AI

Stefan & Jockum

How is the European principled approach to AI governance understood in the Nordic public administration and what are the implications for ADM deployment?

FROM ETHICAL
PRINCIPLES TO
NORMATIVE
IMAGINARIES

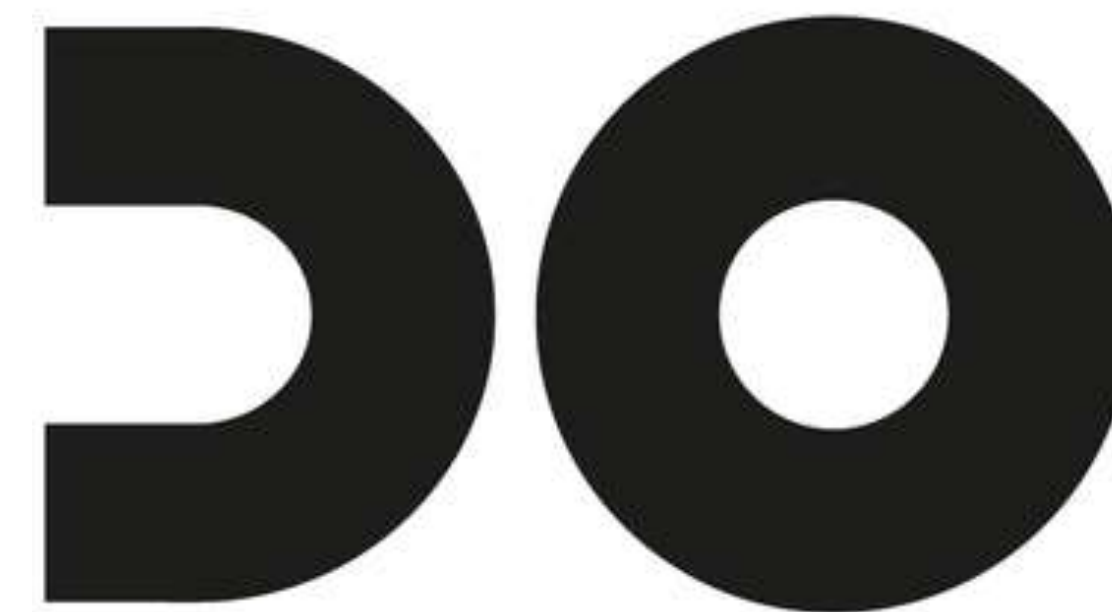
Research to be done:

- Document studies (reports; remittance docs etc.)
- Interview studies, state level authorities
- Coming article: The status of anti discrimination policy in state level authorities' AI developments.

Ongoing Study: AI and Discrimination

Equality Ombudsman (DO)

- DO has the right to information, hence, survey access!
- Discrimination Act (2008:567)
 - Direct discrimination, indirect discrimination, inadequate accessibility, harassment
 - Protected grounds of discrimination: Gender, transgender identity or expression, ethnicity, disability, sexual orientation, age, religion or other beliefs



**Diskriminerings
ombudsmannen**

Survey

- Survey conducted by DO, with advise from me and Charlotte Högberg
- 34 national authorities was asked, 33 answered
- 20 questions on
 - the use of ADM, AI, profiling
 - presence of internal policies on prevention of disadvantage, grounds for discrimination, risk analyses and quality monitoring
 - if cases on ADM and discrimination had been discovered; if they saw a need for development in the field;
 - if individuals are informed



What discrimination perspective? (1/2)

- 10 state that they are doing risk assessments to ensure that individuals are not disadvantaged when implementing ADM
- **But**, only 4 consider the grounds for discrimination
- Many express a need for more competence regarding how AI works and increased knowledge of IT in general
- **But**, only a few consider that they have a need to develop their work to decrease the risk of discrimination with ADM



What discrimination perspective? (2/2)

- Active discussions on ethics, integrity and GDPR
- **But**, the authorities largely lack a discrimination perspective
- DO: Knowledge of the risks of discrimination and obstacles to equal rights is not satisfactory



Tentative sum

1. **Swift formative period of guidelines:** “Strong on principles, weak on implementation”
2. **There is a normative influence on states and organisations:**
But not necessarily in law.
3. **Core ideas:** Transparency, responsibility/accountability, fairness — but somewhat unclear meanings.
4. **ADM-Gov study begins** — Impact and practices of state level authorities? Laws coming, how will they be implemented?



Thank you!

Stefan Larsson

Associate Professor in
Technology and Social Change,
Department of Technology and
Society, LTH, Lund University

Lawyer (LLM)
PhD in sociology of law
PhD in spatial planning fysisk planering



LUNDS
UNIVERSITET