Trustworthy AI
Towards Actionable AI Ethics and “Better” AI Regulation

#EBDVF19
Evolution usage of personal digital assistants, Source: Tractica 2018
What is essentially human in the age of AI?
Timo Honkela (Professor, University of Helsinki)

European Commission’s perspective on ethics in AI
Juha Heikkilä (Head of Unit, DG CNECT A1. Robotics and Artificial Intelligence, EC)

Identification and mitigation of biases in AI based systems.
Practical results
Alexander Benítez Buenache (GMV)

Operationalization of AI ethics: guidelines and good practices
& The opportunities of regulatory sandboxes
Sonja Zillner (Senior Key Expert, Siemens)
Marina Da Bormida (R&I Lawyer and Ethics Expert)
What is essentially human in the age of AI?

Timo Honkela (Professor, University of Helsinki)
European Commission’s perspective on ethics in AI

Juha Heikkilä (Head of Unit, DG CNECT A1. Robotics and Artificial Intelligence, EC)
Identification and mitigation of biases in AI based systems. Practical results

Alexander Benítez Buenache (GMV)
1. WHAT IS FAIRNESS ML?
2. HOW TO IMPLEMENT IT?
3. PRACTICAL RESULTS
4. CONCLUSIONS
WHAT IS FAIRNESS ML?

BEST PRACTICES MITIGATING BIAS
FAIRNESS ML

Allows to design equitable (fair) predictive models to obtain non-discriminatory decisions
WHAT IS FAIRNESS ML? AN EXAMPLE

Biased historical data:
- Male: NO, YES, YES, YES
- Female: YES, YES, NO, YES

Biased models:
- Male: YES
- Female: NO
BEST PRACTICES MITIGATING BIAS

HOW TO IMPLEMENT IT?
HOW TO IMPLEMENT IT?

Best Practices Mitigating Bias

Biased data

- Correlation among variables
- Protected variable is implicit in other ones

Remove the protected variable

It is not enough
HOW TO IMPLEMENT IT?

1. Bias detection in data

2. Bias detection in ML models

3. Mitigation

**Fairness Measures:**
- Statistical Parity Difference
- Disparate impact
- Equal Opportunity Difference
- Average Odds Difference

**BEST PRACTICES MITIGATING BIAS**
- Pre-processing
- Training
- Post-processing
PRE-PROCESSING APPROACHES:

**Uniform sampling:**

- Methodology: Joining samples by subsets + Re-sampling (weighted)

<table>
<thead>
<tr>
<th></th>
<th>Privileged class</th>
<th>Unprivileged class</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Favorable target</strong></td>
<td>Under-sampling (remove)</td>
<td>Over-sampling (duplicate)</td>
</tr>
<tr>
<td><strong>Unfavorable target</strong></td>
<td>Over-sampling (duplicate)</td>
<td>Under-sampling (remove)</td>
</tr>
</tbody>
</table>

- Variations: k-NN sampling based on SMOTE over-sampling technique (GMV)
IN-PROCESSING APPROACH:

Adversarial Debiasing:

- Methodology: Two learning machines (predictor and adversary) sharing information to mitigate the bias effect
PRACTICAL RESULTS

Net Salary Distribution per Genre (2010)

- Female distribution
- Male distribution

Net Salary (grouped in the order of hundreds)
BEST PRACTICES MITIGATING BIAS

PRACTICAL RESULTS

- Bias detection and mitigation in the dataset

**Statistical Parity Difference**

- **Ideal**: 0
- **Biased**: < -0.1

**Disparate Impact**

- **Ideal**: 1
- **Biased**: < 0.8
BEST PRACTICES MITIGATING BIAS

PRACTICAL RESULTS

- Bias detection and mitigation on the predictions made by the model
BEST PRACTICES MITIGATING BIAS

PRACTICAL RESULTS

Performance

Accuracy

Balanced Accuracy

Performance / Bias Trade-off

- Original
- Unif. Sampling
- KNN Sampling
If we allow that *machines* take decision by humans, at least we have to ensure that these decisions are *equitable and non-discriminatory*
THANKS

Alexander Benítez-Buenache
abenitez@gmv.com
Operationalization of AI ethics: guidelines and good practices & The opportunities of regulatory sandboxes

Sonja Zillner (Senior Key Expert, Siemens)
How to operationalise Ethical AI in Industry?

EBDVF2019 in Helsinki 14.10.2019

Dr. Sonja Zillner, Siemens AG, Corporate Technology
Digitalization and AI is disrupting entire customer value chains

Enabling the next level of ...

- ... productivity and time-to-market ...
- ... flexibility and resilience ...
- ... availability and efficiency ...

Design and engineering

Automation and operation

Maintenance and services

Data analytics
Artificial Intelligence
Simulation tools
Cloud & platform technology
Secure connectivity
Cyber-Security
Where is Ethical AI needed?
Which AI applications have significant ethical implications?

Distinguish between AI applications that are solely technical versus those that involve human interaction.

Ethics should be considered in all AI Applications.

AI applications with **human interaction** require significant **ethical consideration**.
Improved efficiency
Sensing & Connectivity & Learning & Acting

- More than 200 GB of sensor data from ≈ 7,800 wind parks
- Use of Reinforcement Learning
- Early detection of divergent behavior
- 1-3% increase of annual energy harvest

Common research project ALICE: Siemens, IdaLab GmbH, TU Berlin
More than 99.9999991% the speed of light and a 7,000,000,000,000,000 proton-proton collision in 2016.

Huge supervisory system and hundreds of controllers supporting the operation.

With rule and pattern mining methods increase operating hours.

Source: CERN
Supporting radiologists in their decision making process by improved visualization of radiological findings. Radiological findings are automatically extracted by means of machine learning and semantics.

**But:** How can we ensure that all **relevant patient’s data** are considered, and whom will you trust your data?
Knowing how many people are in a room or a building is highly relevant.

It may result in additional user comfort and a significant reduction of energy cost.

But: Privacy-sensitive data (e.g. videos) may be needed to get required information.

Why not: develop and make use of technology as trustee (e.g. AI on the edge).
Regulation versus Innovation
Artificial Intelligence is a transversal technology......

......that is applied differently within specific sectors and application scenarios.....

1. There is **only limited scope for a one-size-fits all approach** to AI policies or regulation
2. Regulatory frameworks will need to be adapted or tailored to the **specific sector requirements** or usage scenarios
3. **Horizontal principle-based frameworks** should be complemented by appropriate legal and regulatory amendments in different vertical domains
4. Establish **regulatory sandboxes** as safe environments where novel AI technologies and regulatory solutions can be tested, improved, & verified
Responsibility
Responsibility

The **machine** must be designed to be responsible

Responsibility must be safeguarded by **humans**

This refers to ........

....the capability of the AI systems
....the system’s decisions and its non-biased usage of data

....the role of people interacting with it
....the actions of stakeholders involved in the system’s decision.

Responsibility must be linked to the people developing and using the AI system.
Responsible AI @Siemens
Leveraging AI benefits by reducing risks through clear responsibility

1. AI comes along with plenty of capabilities suited to address several key business needs

2. They create several benefits…

3. …but may also cause risks

4. Responsible use of technology can help to both unfold the benefits and mitigate the risks

- Most mitigation fields can be tackled with existing set of regulations
- Technology & domain know-how needed for implementation

---

<table>
<thead>
<tr>
<th>Capabilities</th>
<th>Benefits</th>
<th>Risks</th>
<th>Mitigation fields</th>
</tr>
</thead>
</table>

Unrestricted © Siemens AG 2019
Page 39 14. October 2019

Sonja Zillner – EBDVF2019
Summary & Conclusion
## Summary: How to operationalise Ethical AI in Industry?

### “Areas of focus"
- Distinguish between AI applications that are solely technical in nature and those that involve human interaction

### “Regulation versus innovation"
- AI is applied differently in specific sectors: horizontal principles should be complemented with vertical regulations
- Regulatory sandboxes are promising to explore

### “Responsibility"
- Provide horizontal principles that support engineers and user in developing ethical-aware AI applications (beside technical security and safety issues)

---

**Ethical AI relies on successful collaboration between multiple disciplines**
Thanks for your attention!

Questions?

Siemens Corporate Technology - Business Analytics and Monitoring

200 Data Scientists & AI experts at 9 locations globally

Dr. Sonja Zillner

Siemens AG
Corporate Technology
Munich, Germany

sonja.zillner@siemens.com

siemens.com
Operationalization of AI ethics: guidelines and good practices
& The opportunities of regulatory sandboxes

Marina Da Bormida (R&I Lawyer and Ethics Expert)
Operationalising AI Ethics: challenges and opportunities from a SMEs’ perspective

Marina Da Bormida
R&I Legal Advisor and Ethics Expert

“BDVF2019, Helsinki, 14th October 2019
SMEs’ viewpoint

**Brainstorming & informal feedback collection**

**Limited number of SMEs** from different domains and background

Free-flowing conversation through **wide and open questions**
Opportunities

- Some of AI ethics requirements are seen as aligned with streamline tasks like auditing and risk assessment.
- Ethical practices as a possible competitive factor (but not full awareness).
- Familiarity with ethical data collection and processing.
- Ad-hoc compliance supporting services?
Barriers and practical issues/1

**Foreword:** AI itself mostly perceived as an **expensive and complex solution** and even more data ethics

Lack of perception of the **future direct benefits** and **unnecessary**

- Lack of long-term vision
- Cultural resistance

**Limited resources**

- **Time and cost** constraints
- Perception of **irresponsible use** of time and resources
- **Disproportion** between necessary efforts (hard work and research) and available resources
Barriers and practical issues/2

- Far from their **daily business work** and customer base’ demand: gap
- Lack of **data ethics talent** (no specialised staff in ethics issues)
- **Limited knowledge**, information and awareness
- **GDPR-centric vision**
### 1. Human agency and oversight

- **Unfamiliar with most topics** (fundamental rights impact assessment, trade-offs, ...)
- Importance of **human oversight & control** and human capabilities’ **enhancement** (stop button or procedure)

### 2. Technical robustness and safety

- Comfortable with concepts like **integrity and resilience** against potential attacks and less with others like **dual use**
- **Risk and safety assessment practices** already in place
- **Insurance policy** as a good idea but cost issues
- Difficult to understand the practical implications of **liability and consumer protection rules** (lack of knowledge)
- Awareness of the link between **data accuracy** and the risk of inaccurate predictions
Focusing on requirements / 2

3. Privacy and data governance

- **Comfortable with many indications** (sensitive or personal data, notice and control, PETs, restricted access,...)
- Already in place efforts for alignment with relevant **standards** (ISO, IEEE) or widely adopted **protocols** for daily data management and governance
- Scarce (or lack) of control of the **quality of the external data sources**

4. Transparency

- **Limited perception of the usefulness and reasons** behind the indications on traceability, explainability and communications (ex. disclaimer for communicating to end-users the interaction with an AI system)
- **Protection of intangible assets**
5. Diversity, non discrimination and fairness

- Agreement on the need to avoid unfair bias in the AI system (input data & algorithm design)
- Most of indications perceived as time & cost consuming and far from daily business (such as processes to test and monitor for potential biases, working definition of “fairness” to be applied in designing AI system)
- In principle, importance of considering vulnerable groups (such as people with special needs or disabilities) and stakeholders’ participation, but practical difficulties
6. Societal and environmental well being

- Importance of environmental and societal impact but no particular insights on that
- Perceived as far from their interests and control (ex. type of energy used by the data centres)?

7. Accountability

- training and education and ethics boards sound promising but not adequate to their dimensions, if implemented at company level
- not easily feasible neither to involve third parties (e.g. suppliers, consumers, distributors/vendors), nor to organize redress mechanisms in case of harm
The way forward...reflecting on facilitating measures

Synergy with IEEE Global Initiative for on Ethics of Autonomous and Intelligent Systems (A/IS): “from principles to practice”

Share AI resources produced in European projects, including high-level services (European Ethical Observatory)

Cascade funding for supporting SMEs in benefitting from AI resources available on the platform and solving AI challenges
The way forward...reflecting on facilitating measures/ 2

**Regulatory Sandboxes**

- **Fintech universe**
- **Small scale, live testing of innovations in a controlled environment**
- **Main features: special exemption & regulator’s supervision**
- **More open and active dialogue between regulators and innovators**
- **Revise and shape the regulatory framework with agility**
- **Extension of DIHs’ function as experimentation facilities? Clear framework needed**
The way forward...reflecting on facilitating measures / 3

- **Incentives** and awareness campaign
- Participation to **European Projects**
- Addressing the value-chain’s **asymmetries of power**
- Business ecosystem services (such as **DIHs’ services**) at EU level (such as training, certification)
Thank you!

Marina Da Bormida
R&I Legal Advisor and Ethics Expert
m.dabormida@eurolawyer.it
+393498433690

“EBDV2019”, Helsinki, 14 October 2019
Thank you all! Let’s continue the debate!