



New skills to innovate: Checks and balances to the potential of AI (intended and unintended consequences)

Jonathan Armstrong
Cordery

Alejandro Bes
Novartis

Introduction

- What is AI?
- Machine Learning vs. AI
- GIGO (e.g. ICO focus on inputs & outputs)

“At present, AI is far more ‘artificial’ than it is ‘intelligent’”

Gregory Stern, Global Integration Counsel, Chubb

A quick case study...



**New skills to innovate: Checks and balances to
the potential of AI (intended and unintended
consequences)**

Alejandro Bes

Areas of Interest

- Practical Principles
- Industry applications
- Risk & Issues
- Mitigating measures
- How training influences output
- What worked – Industry applications



Pixabay licensed – Free for commercial use

Industry applications

“The malicious use of AI will impact how we construct and manage our digital infrastructure as well as how we design and distribute AI systems. “

Miles Brundage et al.

The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation.

arXiv:1802.07228

General ethical principles on AI systems

- 1. Human
- 2. Well-being
- 3. Data Agency
- 4. Effectiveness
- 5. Transparency
- 6. Accountability
- 7. Awareness of Misuse
- 8. Competence

• *Source: IEEE , General Principles of Ethically Aligned Design*

Risks & Issues

- Data Protection & Privacy – improper use of data
- Safety & Security – Reliability of outputs, safe and secure through operational lifetime
- Openness – Fights against the black box effect and failure transparency
- Bias
 - Voluntary vs. Involuntary – Incomplete data
 - Problems derived from the interpretation and inferences about outputs
 - Anchoring effect
- Decision making – Shift of power about the ultimate decision maker

How training influences output

Machine Learning

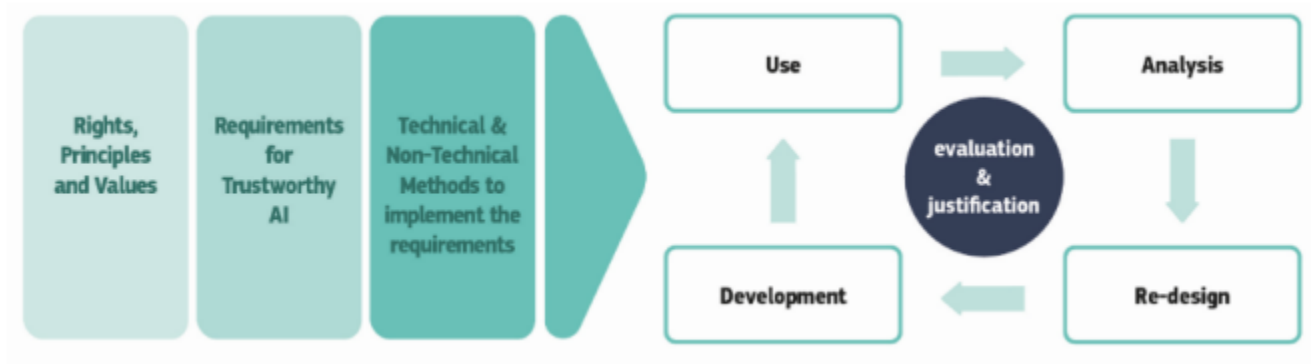


XKCD. CC 2.5 License
<https://xkcd.com/1838/>

Industry examples

- Health and well being – smarter and more targeted treatments, improved analysis and diagnosis and patient support
- Workplace recruitment – targeted and accurate selection of candidates, workplace support and improved underrepresented population hiring practices
- Finance – Improved financing processes and customer relationship
- Transportation – Improved processes and systems. Transition from semi autonomous to fully autonomous AI driving.

Lifecycle effort



Ethics Guidelines for Trustworthy AI,
High Level Expert Group on AI

Suggested readings

- ***Ethically Aligned Design, First Edition***
 - The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems
- **The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation**
 - Brundage et al. arXiv:1802.07228
- **Building Trust in Human Centric AI**
 - <https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines#Top>

Applications

- Be clear about AI vs. Machine Learning
- Don't bake in failure
- Be self-analytical
- “Train your dragon wisely”
- Devote appropriate time and seniority

Risks

- Psychology - Man vs. Machine
- Big Data vs. Data Protection
- Data Ubiquity vs. Data Security

GDPR 6 Principles

1. Lawfulness, fairness and transparency.
2. Purpose limitation – make sure you only collect data for specified, explicit and legitimate purposes.
3. Data minimisation – make sure your use of data is adequate, relevant and limited to what is necessary.
4. Accuracy – make sure your data is accurate and, where necessary, kept up to date.
5. Storage limitation – don't keep data for longer than necessary.
6. Integrity and confidentiality – make sure data is processed in a manner that ensures the appropriate security of personal data, including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage. Make sure appropriate TOMs are in place.

Mitigation

- DPIA / detailed risk analysis (mandatory under GDPR)
- Hambach Declaration - – include non-GDPR legal risks e.g. discrimination
- UX and psychology
- Proper legal agreements (including 6 principles, ground truth assessment etc.)
- Know your data – the data you created + the data your tool creates
- Have a proper DSR strategy (including SARs, RTBF, rectification)

Resources

- GDPR FAQs – <http://bit.ly/eugdprfaq>
- Breach Navigator - <http://bit.ly/breachnav>
- ICO May 2019 Guidance - <http://bit.ly/2GXUV73>
- Judge's task force - <http://bit.ly/2JdHYII>
- Singapore Infocomm Media Development Authority framework - <http://bit.ly/2GTdskl>

Contact



Jonathan Armstrong

jonathan.armstrong@corderycompliance.com

+44 (0) 20 7118 2700

Twitter: @armstrongjp

Questions and Answer Session

Jonathan Armstrong
Cordery

jonathan.armstrong@corderycompliance.com

+44(0) 20 7075 1784

Alejandro Bes
Novartis

alejandro.bes@novartis.com