ISO / IEC AI Workshop Series

Inaugural Workshop

Wael William Diab
ISO/IEC JTC 1/SC 42 Chair
ISO/IEC AI Workshop Chair

SC 42 – Artificial Intelligence
Welcome
Thank you for participating!

Participation Guidelines

Please use the Q&A function for questions

Please do not use the chat function for questions

This webinar is being recorded and will made available

Please keep your video off and audio muted unless speaking
Acknowledgments

Program Committee
- Heather Benko
- Norbert Bensalem
- Andrew Dryden
- Peter Deussen
- Rohit Israni
- Mike Mullane
- Catherine Nelson
- Wael William Diab

ISO
- Elisabeth Gasiorowski
- Catherine Infante
- Vivienne Rojas

IEC
- Stephen Dutnall
- Gabriella Ehrlich
- Ian Gardner
- Giulia Pizzi
- Laura Rocha
Program

SESSION 1
(24th May 13:00 – 16:00 UTC)
- Opening Remarks
- AI Applications
- Novel AI Standardization

SESSION 2
(25th May 05:00 – 08:00 UTC)
- Opening Remarks
- Emerging AI Requirements
- Emerging AI Technology Trends

SESSION 3
(25th May 22:00 – 26th 01:00 UTC)
- AI Applications
- Novel AI Standardization
- Closing Remarks
### Session 1

<table>
<thead>
<tr>
<th>Program</th>
<th>Talk Title</th>
<th>Track Chairs</th>
<th>Aprox Start Time (UTC)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Kickoff</strong></td>
<td></td>
<td></td>
<td>13:00</td>
</tr>
<tr>
<td>Wael William Diab</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Opening Remarks</strong></td>
<td></td>
<td></td>
<td>13:05</td>
</tr>
<tr>
<td>Gilles Thonet</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phil Wennblom</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>AI Applications</strong></td>
<td></td>
<td>Rohit Israni Catherine Nelson</td>
<td>13:15</td>
</tr>
<tr>
<td>Mike Glickman</td>
<td>Health informatics and AI, the road to Interoperability</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neil Frost</td>
<td>AI potential Applications in Transport</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Christophe Preube</td>
<td>AI finds awareness in standardization work – the White paper of ISO/TMB SMCC</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Novel AI Standardization Approaches</strong></td>
<td></td>
<td>Peter Deussen Norbert Bensalem</td>
<td>14:25</td>
</tr>
<tr>
<td>Kimberly Lucy</td>
<td>Creating Trust in AI Through Standards: A Management System Approach</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wo Chang</td>
<td>Data Quality for Analytics and Machine Learning</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tim McGarr / Florian Ostmann</td>
<td>AI Standards Hub (UK)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Viveka Bonde</td>
<td>Novel Standardization Approach to AI Ethics</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Session 2

<table>
<thead>
<tr>
<th>Program</th>
<th>Talk Title</th>
<th>Track Chairs</th>
<th>Aprox Start Time (UTC)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Kickoff</strong></td>
<td></td>
<td>N/A</td>
<td>5:00</td>
</tr>
<tr>
<td>Wael William Diab</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Opening Remarks</strong></td>
<td></td>
<td>N/A</td>
<td>5:05</td>
</tr>
<tr>
<td>Silvio Dulinsky</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Emerging AI Requirements</strong></td>
<td></td>
<td>Catherine Nelson and Peter Deussen</td>
<td>5:10</td>
</tr>
<tr>
<td>Elham Tabassi</td>
<td>NIST AI Risk Management Framework</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mariagrazia Squicciarini</td>
<td>The UNESCO Recommendation on the Ethics of AI - Setting the standards for a better and more inclusive future</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Liz Coll</td>
<td>Can standards deliver consumer trust and confidence in AI</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Daniel Loevenich</td>
<td>Evaluation Standards for Conformity Assessment of Trustworthy Cloud-based AI Applications</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Emerging AI Technology Trends</strong></td>
<td>Emerging AI Trends – as seen by an Industry Practitioner</td>
<td>Norbert Bensalem and Rohit Israni</td>
<td>6:30</td>
</tr>
<tr>
<td>Shubhashis Sengupta</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tilak Kasturi</td>
<td>How specialized AI drives value for Automotive Service Organizations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>William Uppington</td>
<td>AI Quality: The Next Big Challenge in AI</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Babak Hodjat</td>
<td>From Data to Decisions, and Back</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Session 3

<table>
<thead>
<tr>
<th>Program</th>
<th>Talk Title</th>
<th>Track Chairs</th>
<th>Aprox Start Time (UTC)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Kickoff</strong></td>
<td></td>
<td>Wael William Diab</td>
<td>22:00</td>
</tr>
<tr>
<td><strong>AI Applications</strong></td>
<td>Safety considerations in autonomous products</td>
<td>Rohit Israni and Catherine Nelson</td>
<td>22:05</td>
</tr>
<tr>
<td>Charalambos Freed</td>
<td>Development of SRD 63416 “Ethical Considerations of AI when Applied in the AAL Context”</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hajime Yamada</td>
<td>AI Powered UAV and Future Prospects</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kenzo Nonami</td>
<td>Use cases and AI application guidelines in international standardization</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fumihiro Maruyama</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Novel AI Standardization Approaches</strong></td>
<td>AI Standardisation supporting regulatory needs</td>
<td>Norbert Bensalem and Peter Deussen</td>
<td>23:30</td>
</tr>
<tr>
<td>Jochen Friedrich</td>
<td>Developing a TS on assessment of machine learning classification performance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linzhing Meng / Mike Thieme</td>
<td>Introduction of Governance Implications of AI Systems</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yonosuke Harada</td>
<td>The foundational standards for AI – ISO/IEC 22989 and ISO/IEC 23053</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paul Cotton / Milan Patel / Wei Wei</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Closing Remarks</strong></td>
<td>Overview and introduction of SC 42</td>
<td>Wael William Diab</td>
<td>0:55</td>
</tr>
<tr>
<td>Wael William Diab</td>
<td>Insights from the workshop</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Program Committee</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Thank you

Wael William Diab

wael.diab@gmail.com; Wael’s LinkedIn

Workshop Website

ISO/IEC AI LinkedIn Page
NIST is actively engaging stakeholders to develop a risk management framework to map, measure, manage, and govern risks associated with AI technologies.
NIST AI Program

- Conduct foundational research to advance trustworthy AI technologies
- Advance AI research and innovation across the NIST laboratory programs
- Participate and lead in the development of standards to advance AI innovation
- Contribute NIST’s technical expertise to discussions and development of policies
- Establish benchmarks and develop metrics to evaluate AI technologies
- Ensure that NIST has resources and expertise to carry out its AI programs
Open and public data

“Yet another advice: don’t get fooled by people who claim to have a solution to Artificial General Intelligence. Ask them what error rate they get on **MNIST** or ImageNet.”

Yann LeCun, 2014.
Director of AI Research, Facebook and NYU professor

NIST Special Database 19 NIST Handprinted Forms and Characters Database

www.nist.gov/srd/nist-special-database-19

www.reddit.com/r/MachineLearning/comments/25lnbt/ama_yann_lecun/
Advance scientific disciplines via evaluations

“Our prior QA work took shape in the form of a QA system called PIQUANT. PIQUANT development started in 1999, predating our work in UIMA, and was funded by government research grants and tested against NIST evaluation data in the Text REtrieval Conference (TREC) QA track between 1999 and 2005.”

Introduction to “This is Watson”. IBM Journal of Research and Development (Volume: 56, Issue: 3.4, May-June 2012)
Cultivate trust in the design, development, use and governance of artificial intelligence technologies and systems.
Trustworthy AI’s Foundation
From Technical Requirements to Policy Creation

- Identify Building Blocks or Technical Requirements
- Concepts, Terminology, and Taxonomy
- Metrics, Evaluations, and Benchmarks
- Risk Management
- Governance
- Policy Considerations
<table>
<thead>
<tr>
<th>AI RMF Taxonomy</th>
<th>Technical Design Characteristics</th>
<th>Socio-Technical Characteristics</th>
<th>Guiding Principles Contributing to Trustworthiness</th>
</tr>
</thead>
<tbody>
<tr>
<td>OECD</td>
<td>• Robustness • Security</td>
<td>• Safety • Explainability</td>
<td>• Traceability to human values • Transparency and responsible disclosure • Accountability</td>
</tr>
<tr>
<td>EU AI Act</td>
<td>• Technical robustness</td>
<td>• Safety • Privacy • Non-discrimination</td>
<td>• Human agency and oversight • Data governance • Transparency • Diversity and fairness • Environmental &amp; societal well-being • Accountability</td>
</tr>
<tr>
<td>EO 13960</td>
<td>• Purposeful and performance-driven • Accurate, reliable, and effective • Secure and resilient</td>
<td>• Safe • Understandable by subject matter experts, users, and others, as appropriate</td>
<td>• Lawful and respectful of our Nation’s values • Responsible and traceable • Regularly monitored • Transparent • Accountable</td>
</tr>
</tbody>
</table>
Trustworthy AI: Risks & Characteristics

**Technical**
- Accuracy
- Reliability
- Robustness
- Resilience or Security

**Socio Technical**
- Explainability
- Interpretability
- Privacy
- Safety
- Managing Bias

**Guiding Principles**
- Fairness
- Accountability
- Transparency
AI Risk Management Framework

Map
Context is recognized and risks related to the context are identified

Govern
A culture of risk management is cultivated and present

Measure
Identified risks are assessed, analyzed, or tracked

Manage
Risks are prioritized and acted upon based on a projected impact
The Challenge of AI Bias

Special Publication 1270 - Towards a Standard for Identifying and Managing Bias in Artificial Intelligence
Four Principles of Explainable AI

1. Explanation
2. Meaningful
3. Explanation Accuracy
4. Knowledge Limits

NISTIR 8312: Four Principles of Explainable Artificial Intelligence

NISTIR 8367: Psychological Foundations of Explainability and Interpretability in AI
Secure AI

A Taxonomy and Terminology of Adversarial Attacks and Defenses on Machine Learning
“If you cannot measure it, you cannot improve it.”

Devise and revise metrics
Testbeds and test data
Assessing real world performance
Interoperable tests and test results
USG AI Standards Coordinator

Outreach to connect with all known federal efforts relating to AI standards development and use with the goal of community participants leveraging and learning from the successes of other participants.

Facilitate ongoing discussions between the U.S. private sector and federal agencies to strengthen private-public sector coordinator
From Innovation to Adoption

Fundamental Research

Applied Research

Standards + Best Practice Guides

Adoption

Image Credit: wsj.com
THANK YOU

Contact us via email at aiframework@nist.gov

For more info on the NIST AI RMF, visit https://www.nist.gov/itl/ai-risk-management-framework
The UNESCO Recommendation on the Ethics of AI

Setting the standards for a better and more inclusive future

Mariagrazia Squicciarini, PhD
Chief of Executive Office
UNESCO Social and Human Sciences Sector
25 May 2022
AI research and development is highly concentrated. The large majority AI patents are filed in and granted to groups in the U.S. and China.
AI HIRING IS GROWING

RELATIVE AI HIRING INDEX by GEOGRAPHIC AREA, 2021

Source: LinkedIn, 2021 | Chart: 2022 AI Index Report

SHARE OF AI-RELATED JOBS IN BGT, BY COUNTRY AND YEAR

Source: Nachtigall and Squicciarini, 2021
GLOBAL PRIVATE INVESTMENT in AI by FOCUS AREA, 2019 vs 2020

Source: CapIQ, Crunchbase, and NetBase Quid, 2020 | Chart: 2021 AI Index Report

- Drugs, Cancer, Molecular, Drug Discovery
- Autonomous Vehicles, Fleet, Autonomous driving, Road
- Students, Courses, Edtech, English language
- Open Source, Compute, Hadoop, Devops
- Speech Recognition, Computer Interaction, Dialogue, Machine translation
- Money Laundering, Anti-fraud, Fraud Detection, Fraud Prevention
- Fashion, Shopping Experience, Beauty, Visual Search
- Games, Fans, Gaming, Football
- Semiconductor, Chip, Data Centers, Processor
- Bank, Card, Credit Cards, Gift

Total Investment (in Millions of U.S. Dollars)

2019 vs 2020
More and more public and private decisions are taken by artificial intelligence, and the outcomes are not free from biases or badly-defined frameworks. Privacy, surveillance, and serious human rights concerns.
**SOCIAL AND HUMAN SCIENCES**

**DOWNSIDES OF SELF-REGULATION**

- Human Rights Violations
- Privacy Issues
- Diversity Inclusion
- Bias
- and more…
LACK OF A GLOBAL STANDARD AND BENCHMARKS
LACK OF A GLOBAL STANDARD AND BENCHMARKS
THE FRAMEWORK

The Recommendation is based on interconnected values and principles

**Values**

1. Respect, protection and promotion of human rights and fundamental freedoms and human dignity
2. Environment and ecosystem flourishing
3. Ensuring diversity and inclusiveness
4. Living in peaceful, just and interconnected societies

**Principles**

1. Proportionality and do no harm
2. Safety and security
3. Fairness and non-discrimination
4. Sustainability
5. Right to privacy, and data protection
6. Human oversight and determination
7. Transparency and explainability
8. Responsibility and accountability
9. Awareness and literacy
10. Multi-stakeholder and adaptive governance and collaboration
Principles were there for many years, but the Recommendation goes into details of actions to ensure accountability, responsibility, transparency and necessary regulations to ensure the rule of law.

The Recommendation includes several action-oriented cross-sectoral policy chapters covering:

Data policy, Gender, Development & international cooperation, Environment & ecosystems, Health and social well-being, Communication & information, Education & research, Economy & Labour, Culture
92% of businesses believe that high-quality data is the fuel for digital transformation.

75% of businesses say that poor quality data has made it challenging to achieve their digital transformation plans.

Only 15% of respondents believe that their existing systems are capable of producing clean data that can be trusted.

Source: 2018 Study by the Supply Chain Resource Cooperative at North Carolina State University and IBM Watson
GENDER

300 millions
fewer women than men access mobile internet

25 %
of all tech jobs are held by women

Women are TWICE as likely as men to have lost their jobs due to the pandemic

3 %
of female students in higher education choose ICT courses
FINANCING GENDER RELATED SCHEMES

INCLUDING GENDER ACTION PLAN

ADDRESSING WAGE AND EQUAL OPPORTUNITIES GAP

ENCOURAGING FEMALE ENGAGEMENT IN AI

INCREASING FEMALE PARTICIPATION IN STEM AND ICT

ERADICATING GENDER STEREOTYPING AND BIAS
EDUCATION AND SKILLS

- Top 30 skills appearing in AI-related jobs tend to be rather “general”
- Many of them relate to the “tools of the job”, i.e. are software-related (e.g. MATLAB, R, SAS)
- Some technical skill bundles emerge
- A number of additional cognitive and socio-emotional skills are also required

Source: Samek, Squicciarini and Cammeraat (2021)
ASSISTING MEMBER STATES TO IMPLEMENT THE RECOMMENDATION

- Ethical Impact Assessment (EIA)
- Readiness Assessment Methodology
- Building Institutional Capacity
- Development of knowledge products

Social and Human Sciences
ASSISTING MEMBER STATES TO IMPLEMENT THE RECOMMENDATION

Observatory of Ethical AI
Global Forum on Ethics of AI
Women4EthicalAI
AI Experts without borders
Analytical works
Thank you!
m.squicciarini@unesco.org

More on the UNESCO Recommendation on the Ethics of AI:
https://en.unesco.org/artificial-intelligence/ethics
Emerging AI requirements

How can standards deliver consumer trust and confidence in AI?

Liz Coll, BSI Consumer and Public Interest Network Steering Group, and UK representative to the ISO COPOLCO AI task group.

Bsigroup.com/consumers
Outline

- Importance of standards for consumers
- Consumer stakeholder engagement in standards
- ISO COPOLCO
Importance of standards for consumers
Importance of standards for consumers

- Consumers **require protection** from: unsafe products, poor service, unfair treatment etc

- Standards are part of the ‘**consumer protection toolkit**’ along with: legislation, regulation, industry codes, enforcement, information, advice

- **Minimize** deliberate or accidental detriment – could be physical, financial, psychological, wider economic/environmental impact

- **Maximise** consumer rights: information, education, redress, fair choice, safety, healthy environment, privacy and security, access, inclusivity

  Meeting consumer protection principles **builds trust and confidence** in products and providers
Consumer stakeholder engagement in standards
The consumer stakeholder

- Representative of an independent consumer organization, working in the interest of consumer and wider public interests
- Consumers International, ANEC, national consumer organisations, independent NSB committees.
- Consumer participation in standards adds value by:
  - Ensuring standards address real problems and experiences faced by real people – crucial for technology standards
  - Technical content is based on relevant data from sources outside of usual scope
  - Products and services are fit for purpose, safer, fairer, better quality, more sustainable, more accessible and inclusive

Minimizes risk of harm and improves confidence
COPOLCO

- COPOLCO is an **ISO policy committees** advising/reporting to the ISO/Council.
- COPOLCO coordinates and encourages **consumer input** into technical and policy work in ISO to help consumers benefit from standardization.
- Provide a **consumers’ network** to exchange information.
- Advise **ISO on policies** and actions to respond to consumers’ needs.
- **Make recommendations** on current and potential standardization work – Privacy by Design, Vulnerable Consumers, Mobile Banking, Cross-border trade of second-hand goods etc.
- Strategic priorities 2022: Sustainability, Digitalization and Vulnerability.
Consumers and Artificial Intelligence
Consumer experience of AI

- Types of tasks AI-enabled technologies perform (in combination or separately)
  - sorting, classifying, pattern recognition, image recognition, voice recognition and natural language processing, predicting events, assessing significance, deciding options, enacting options etc

- Much is labelled as AI when simply an algorithm or when use AI technologies for a portion of their functionality.

- Public understanding is mixed due to jargon and hype

- General purpose vs narrow purpose
Current use cases

• Use cases in action - common in some consumer areas, infancy in others:

  • Voice recognition services (fraud detection/identify verification, verbal chatbots, home assistants)
  • Facial recognition via door entry system (Amazon Ring) or in predictive policing and crowd control (USA)
  • Healthcare or other social care entitlements distribution (USA), sentencing guidelines
  • Recommendations for content, search etc
  • Products/Appliances – energy efficient washing machines etc
  • Recruitment – identification of people from data on their skills, experience, job titles, pay grades etc.
  • Running services – public transit, security systems, traffic management systems, health research etc etc
Is AI different to other technologies?

- AI systems and tools continually learn and adapt their outputs - there is not a static entity to easily observe or test.

- AI technologies are *socio-technical* - operate in human contexts, to fulfil human-defined goals. They reflect the values and choices of those who build and use them (CofE AI, human rights, democracy, and the rule of law, 2021)

- Performance and outcome of an AI product or system does not only depend on its technical components, but on decisions about who uses the technology for what purpose, in what context. (ANEC Position Paper on EU AIA, 2021)

- These decisions have the potential to involve and impact on fundamental rights, cause harm and lower confidence and trust (ANEC)

bsi.
Is AI markedly different to other technologies?

• While AI is different from other sets of technologies, many of the same issues are present—data, security, bias, lack of information, safety, discrimination etc.

• And same consumer rights apply:
  • Right to Transparency, Explanation and Objection
  • Right to Accountability and Control
  • Right to Fairness
  • Right to Non-discrimination
  • Right to Safety and Security
  • Right to Access to Justice
  • Right to Reliability and Robustness

BEUC, AI rights for consumers, 2019
Developing responsible AI

- Developing quickly and being rolled out in markets and public services, without clear governance or regulation.
- Predictions that the use of AI will become more **ubiquitous**. As intro to this track says, it also must be **responsible**.
- What do we need from standards to make sure we develop responsible AI that consumers have trust and confidence in?
Global Risk Perception of AI Decision Making

Will the development of machines or robots that can think and make decisions in the next 20 years mostly cause harm or mostly help?

<table>
<thead>
<tr>
<th>Region</th>
<th>Mostly Harm</th>
<th>Mostly Help</th>
<th>Neither</th>
<th>Do not Know</th>
</tr>
</thead>
<tbody>
<tr>
<td>Latin America &amp; Caribbean</td>
<td>49</td>
<td>26</td>
<td>19</td>
<td>6</td>
</tr>
<tr>
<td>North America</td>
<td>47</td>
<td>41</td>
<td>12</td>
<td>0</td>
</tr>
<tr>
<td>Europe</td>
<td>43</td>
<td>38</td>
<td>15</td>
<td>5</td>
</tr>
<tr>
<td>Central Asia</td>
<td>34</td>
<td>36</td>
<td>17</td>
<td>13</td>
</tr>
<tr>
<td>Middle East</td>
<td>33</td>
<td>38</td>
<td>19</td>
<td>10</td>
</tr>
<tr>
<td>South Asia</td>
<td>33</td>
<td>31</td>
<td>17</td>
<td>19</td>
</tr>
<tr>
<td>Africa</td>
<td>31</td>
<td>41</td>
<td>16</td>
<td>12</td>
</tr>
<tr>
<td>Southeast Asia</td>
<td>25</td>
<td>37</td>
<td>21</td>
<td>17</td>
</tr>
<tr>
<td>East Asia</td>
<td>11</td>
<td>59</td>
<td>12</td>
<td>18</td>
</tr>
</tbody>
</table>

Based on sample of 154,195 respondents in 142 countries
Global Attitudes towards Artificial Intelligence (AI), University of Oxford 2021. Analysis of survey data from the 2019 World Risk Poll
What might erode consumer trust and confidence?

Consumers are positive about potential benefits of AI in some applications but concerns about individual/societal impact. They express fears that:

- it will lead to an increase in abuses related to the use of their personal data
- companies are using AI to manipulate their decisions
- governments are using AI to control their citizens
- AI is dangerous because machines can fail
- AI could lead to unfair discrimination against particular individuals or social categories
- it is not clear who is accountable in case AI is not secure or causes harm

BEUC nine country survey, 2019
Role for standards

• Not all consumer concerns and risks can be dealt with via standards

• Legislation is very important in setting out parameters and expectations (as EU AI Act is poised to do)

• Looking at concerns and risks, there are obvious areas where standards can help eg, data quality, privacy and security, environmental efficiencies, Transparency, redress, design of consumer interfaces for maximum accessibility and understanding
## Potential areas for standardization

<table>
<thead>
<tr>
<th>Design:</th>
<th>Development:</th>
<th>Deployment:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risk assessment</td>
<td>Data quality</td>
<td>Misleading claims</td>
</tr>
<tr>
<td>Impact assessment</td>
<td>Data training</td>
<td>Explainability: visuals, language comprehension etc</td>
</tr>
<tr>
<td>Data procurement, extraction and analysis: reviewing for bias /blanks</td>
<td>Data collection: clarity on consent</td>
<td>Audit: technical documentation and record keeping</td>
</tr>
<tr>
<td>Ethics in design</td>
<td>Model testing: reviewing for bias and negative impact</td>
<td>Supervision</td>
</tr>
<tr>
<td>Stakeholder consultation</td>
<td>Safety</td>
<td>Transparency</td>
</tr>
<tr>
<td></td>
<td>Autonomy</td>
<td>Quality management</td>
</tr>
<tr>
<td></td>
<td>Cybersecurity</td>
<td>Dispute resolution</td>
</tr>
</tbody>
</table>

### Transversal:
- Governance
- Ethics
- Environment and sustainability
- Conformity Assessment
- Terminology
- Classifications
- Privacy and data protection
Consumer requirements within AI standards

- **Risk assessment** standards content to address needs and expectations:
  - Likelihood of the harm occurring
  - Immediacy of the harm
  - Foreseeable use of the AI system
  - Measures and metrics for short and longer term impacts of the AI system on:
    - health, safety, fundamental rights, consumer rights, societal effects, impact on democracy, rule of law or environmental impact, as well as the potential for economic harm.
Consumer requirements within AI standards

• **Inclusion and access** standards content to address needs and expectations:
  - Accessible and usable devices and services for users of AI systems and services
  - Design for all/access to older people and people with disabilities
  - Multi-cultural and multi-linguistic aspects
  - Consumers expect to be able to use products, services and systems enabled by AI whatever their age or ability..
Consumer requirements within AI standards

- **Transparency and information** standards content to address needs and expectations

- AI systems entail very complex technical processes which are hard to explain and understand for consumer users of products and services.
  - Explanations about how decisions are made and how the system works are tailored to the context and provided in useful and comprehensible format suitable for user and include future potential risks of AI.
  - Information on the nature of data use and potential privacy implications for users of AI products and services is clear and readily available.
  - Information is also clearly and readily available on how to object to decisions or seek human recourse.
Consumer requirements within AI standards

- **Cybersecurity** standards content to address needs and expectations:
- The personal information and personally identifiable information in use in AI systems increases the cybersecurity risks.
  - Technical requirements such as secure authentication and security updates, incident management and abuse alerts, encryption etc.
  - Organisational requirements such as information on product security status and support, expected product lifetime etc.
Consumer requirements within AI standards

- **Quality management** standards content to address needs and expectations. A quality management system which includes:
  - Robust consumer relationship management
  - Systems for dispute resolution, complaints
  - Customer redress
  - Effective objection systems for dubious decisions
  - Recall of harmful products
  - Post-marketing monitoring system
Thank you

Liz Coll, BSI Consumer and Public Interest Network Steering Group, UK representative to the ISO COPOLCO AI task group.
Evaluation Standards for Conformity Assessment of Trustworthy Cloud-based AI Applications

Daniel Loevenich, 16. Mai 2022
Agenda

Understanding AI Applications

Understanding ISO/IEC AI Standards

Horizontal Trustworthy AI Conformity Assessment

Project Success Factors

A uniform Framework for Horizontal AI Conformity Assessment
Understanding AI Applications
Understanding AI Applications

- **Conformity Assessment** considers the complex technical system (e.g. a vehicle) as a whole entity in the application context.
- **Risks and requirements for the AI-components** have to be derived from this.
- Such **Operationalizations** are essentially **mappings** into AI requirements.
- This results in evaluation requirements for each component of usually hybrid AI solutions (**application-agnostic** perspective).
- A technical system usually follows existing test regulations for supply chains with multiple actors, e.g. OEMs
- Industry expects cascadable evaluations for (distributed) AI-systems
- AI test results must be composable (for instance in case of individual component checks of hybrid systems)
- A horizontal standard for trustworthy AI must be embeddable into AI management systems (AIMS)
Understanding ISO/IEC AI Standards: Testing and Evaluation

Horizontal and Vertical Standards Concept:

- **Downgrading Risks** to AI Components
- **Conformity Testing and Evaluation** of AI Components (all CA types)
- Upgrading and **Composition** of Evaluation Results
- **Uniform Conformity Assessment Framework**
- Application of schemes and **introduction to markets worldwide**
Horizontal Trustworthy AI Conformity Assessment: Strategic Objectives

Trustworthiness of the entire Supply Chain becomes transparent with Conformity Assessment.

The necessary evaluation criteria and test procedures are to be developed.

This evaluation bases shall be applicable to hybrid solutions as well as embedded components.

**EU-AIA**: The evaluation bases serve as foundation for a horizontal AI standard.

An easy market access for SMEs with acceptable costs will be facilitated.

Vertical and sectoral standards should be based a horizontal standard for trustworthy AI.

**Uniform Evaluation Framework for Trustworthy AI Applications**
Horizontal Trustworthy AI Conformity Assessment: Deliverables

An extendable set of horizontal, application agnostic criteria (“Trustworthy AI Services Evaluation Criteria”),

An extendable set of valid AI evaluation procedures (“Trustworthy AI Services Evaluation Methodology”), applicable to all three types of conformity assessment (self assessment, attestation and certification),
Horizontal Trustworthy AI Conformity Assessment: Deliverables

A proposal for an **application procedure** to extend the methodology and to implement it within the ongoing standardization process,

A proposal for an **application procedure** to extend this criteria and to implement the procedure within the ongoing standardization process,

A **procedure for mappings** of vertical application specific requirements into horizontal criteria requirements,

The **Guidance documents for production, application, and support** for all parties involved in the corresponding AI Eco-Systems. In the context, the framework establishes additional guidelines on how to integrate evaluation activities into an AIMS.
# Horizontal Trustworthy AI Conformity Assessment: Project Roadmap

<table>
<thead>
<tr>
<th>Definition of evaluation bases</th>
<th>Applicability and Market Penetration</th>
<th>Publication Phase</th>
<th>International Standardization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q3 - 2022</td>
<td>Q1 - 2023</td>
<td>Q4 22/Q3 23</td>
<td>2025</td>
</tr>
<tr>
<td>Development of criteria, methodology, scheme with use cases</td>
<td>Validation and Acceptance on basis of relevant Use Cases</td>
<td>Publication within three Workshops - Europe, USA, Asia</td>
<td>Transfer and Harmonization in hEN/ISO Standard</td>
</tr>
</tbody>
</table>

- **Validation** and Acceptance on basis of relevant Use Cases
- Publication within **three Workshops** - Europe, USA, Asia
- Transfer and Harmonization in hEN/ISO **Standard**
Project Success Factors: Drivers for Technology Transfer

Cloud Providers drive AI technologies within the compete AI supply chain. They offer full service enterprise customer support with AI experts for development, IDEs, Frameworks and quality measurement tools and operate AI solutions continuously.

A project for an horizontal standard for trustworthy AI shall use these resources.
# Project Success Factors: Use Case Projects Outline

<table>
<thead>
<tr>
<th>Phase 1:</th>
<th>Phase 2:</th>
<th>Phase 3:</th>
<th>Phase 4:</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Focus:</strong></td>
<td>Scoping, Risk Analysis &amp; TOE</td>
<td>Evaluation Procedure Development</td>
<td>Procedure Execution</td>
</tr>
<tr>
<td><strong>Format:</strong></td>
<td>Workshops</td>
<td>Coordination Conference with unanimous vote</td>
<td>Depends on the type of conformity assessment</td>
</tr>
</tbody>
</table>

**Mandatory:** Generalization of Requirements up to Criteria and Evaluation Process Definition

**Optional:** Pilot Evaluation, Procedure Validation, Assurance Methods Assessment (Levels)
### Project Success Factors: Sub-Projects and Use Cases

#### Modularity and Extendability

<table>
<thead>
<tr>
<th>Prozess</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heath Care</td>
<td>(Mapping) AI Life Cycle: definition, implementation and execution of AI-processes for development and operation</td>
</tr>
<tr>
<td>Financial Services</td>
<td>(Mapping) Different AI-Technologies (e.g. Deep Learning or Reinforcement Learning) cover the whole risk assessment process.</td>
</tr>
<tr>
<td>Automotive</td>
<td>(Mapping)</td>
</tr>
</tbody>
</table>

#### Technology Criteria

- Different AI-Technologies (e.g. Deep Learning or Reinforcement Learning) cover the whole risk assessment process.
A uniform Framework for Horizontal AI Conformity Assessment: End User Acceptance

Open questions for dealing with AI in a trusting manner …

- How can responsible use of AI be demonstrated across the entire supply chain of an AI system?
- Which requirements are relevant for my AI systems and how do I fulfill them according to my area of responsibility?
- How can I efficiently implement requirements with regard to my use case - specific systems and processes?
- How do I efficiently demonstrate compliance with increasing AI regulations?

... are addressed by applying the criteria and method documents

AI Service Providers as important multiplicators

Provision of extendable and horizontal criteria for trustworthy AI applications
Provision of a Validation, Testing and Certification attempt
A uniform Framework for Horizontal AI Conformity Assessment: Standardization System Pyramide

Transfer of AI Trustworthiness Standards
Conformity Assessment of AI-Standard Solutions lead to Acceptance of AI Ecosystems
Worldwide End Customer Support in regulated sectors: Market Penetration

Complete AI Ecosystem Coverage
### Project Partners

#### Development, Validation, Market Penetration

- **Hyperscaler:**
  - AWS
  - Alibaba
  - Google
  - Huawei
  - Microsoft
  - Oracle
  - SAP
  - Dt. Telekom

- **Research Projects:**
  - DFKI
  - Fh IAIS (Program „Certified AI“)
  - With Use Cases: GAIA-X

#### Standardization

- **Evaluation, Certification and Accreditation:**
  - DIN/VDE
  - CEN/CENELEC AI JTC21
  - ISO/IEC JTC 1/SC42

- **AI Quality und Testing Hub**
- **DFKI Osnabrück**
- **Fh IAIS**
- **PricewaterhouseCoopers**
- **Verband der TÜV**
Das BSI als die Cyber-Sicherheitsbehörde des Bundes
gestaltet Informationssicherheit in der Digitalisierung
durch Prävention, Detektion und Reaktion
für Staat, Wirtschaft und Gesellschaft.

Contact:
Daniel Loevenich
daniel.loevenich@bsi.bund.de
Tel. +49 (0) 228 9582 5395
Fax +49 (0) 228 10 9582 5395
Mobil +49 (0)171 30 29 824

Bundesamt für Sicherheit in der Informationstechnik (BSI)
Referat TK 23: Grundsatz, Strategie und Nachweise in der
Künstlichen Intelligenz
Godesberger Allee 185-189
53175 Bonn
www.bsi.bund.de

Thank you for your attention!
Emerging AI Technology Trends

Shubhashis Sengupta

accenture
Introduction

5 Key Technology Innovation Trends are shaping the future of AI engineering and Adoption

- Large Language and Image Learning Models
- Gaining Efficiency and Sustainability in AI
- Synthetic Data comes to the Rescue
- AI and Intelligence
- AI Automation at Scale
TREND1: Large-Scale “Meta-Learning” Models

- Transformer models – learning from Terabytes of data and Billions of Parameters
- 3 key Innovations –
  - Attention
  - Self Attention
  - Positional Embedding
- MIT calls these Foundation Models

- Usage and adaptation are changing dramatically –
  - Few Shot and Zero-Shot learning through “Smart Prompts”
  - Fine Tuning of models through LLRD, SWA, Adaptive Tokenization, LoRA
  - Model access, open sourcing and BERTOLOGY
- Meta Learning as models learn representations for multiple tasks (summarization, translation, Q&A, classification, lang generation, sentiment analysis etc.)

- Similar advancement in Image Language Modeling and Generation –
  - DALL-E, DALL-E2
  - Multi-modal Transformers

Image courtesy: Cross Validated, Stack Exchange

Image courtesy: https://twitter.com/OpenAI/
Image courtesy: lilianweng.github.io

Image courtesy: ISO/IEC AI Workshop | 24 – 25 May 2022
TREND2: Efficient AI (Faster, Cheaper, More Sustainable)

- Large Models are Costly and not Environment friendly
  - Google Switch (1.6 T params) takes 1 MUSD to train and emits 284T of Co2e
  - Training GPT-3 once may take up to 800K USD

- Making AI training and Optimization efficient at Data Centers
  - Algorithmic Optimization – Pruning and Clustering
  - Deployment Optimization – Equalization, Fold-Batch norms, Fused layers, Quantization

- Training speedups: FFT models, Pathway Dataflow for parallelism

- How to induce Pruning / Sparsity without Model Degradation

Trained Models -> Candidate Models -> NetOpt Models -> DepOpt Models

Image Courtesy: Qualcomm

Efficient Edge Computing Architecture
- TinyML models for low AI footprint
- Optimized models like MobileNet V1 (DS-CNN), V2
- Neuromorphic Architecture (SNN) – extremely low power architecture

Table 3: Estimated cost of training a model in terms of CO2 emissions (lbs) and cloud compute cost (USDS). Power and carbon footprint are credited for TPUs due to lack of public information on power draw for this hardware.

Source: https://blog.learningtree.com/carbon-footprint-ai-deep-learning/

Image Courtesy: Qualcomm

Spiking Neural Network: https://en.wikipedia.org/wiki/Spiking_neural_network

https://www.tinyml.org/
TREND3: Synthetic Data for AI

- **Right Data is the most Difficult Ask**
  - Getting right domain data for Model Training is getting increasingly difficult
    - GDPR Regulation, Sensitive (PHI, PII) data, Localization and Fiduciary implications, Data Security

- **Open Source and Linked Data exploitation has hit limits**

- **Data Generation for AI**
  - Data De-identification
  - Data Augmentation
  - Fully Synthetic Generative Modeling
    - Models with Implicit likelihood (GAN -> generates by comparison)
    - Models with Explicit likelihood (VAE, Fully Observable Models)
    - Diffusion Network (very realistic)

- **Quality and Trustworthiness**
  - Should not copy but preserve relationship/correlation
  - Statistically meaningful
  - Predictivity, Diversity, Realism, Privacy Preservation

Image Courtesy: iri.com
Image Courtesy: SecureDrive

Source: ICML Tutorial on Synthetic Data Generation (https://www.youtube.com/watch?v=_EEH9HU2EE0)
TREND4: Cognitive AI (Can AI augment / partner with Humans?)

- Narrow AI supremacy (already achieved)
  - AI is already performing at "superhuman" levels for many specific jobs (Q&A, Reading Comprehension, Image Segmentation)
  - Generalization to multi-task level is the key challenge

- Towards Complex and Creative Tasks (In Progress)
  - Planning and Strategy (Games) – Deep Reinforcement Learning (AlphaZero) vs Rules
  - Common Sense Reasoning – Atomic
  - Analogy based Reasoning, Forward Chaining
  - NLI / NLE
  - Multi-step Reasoning with Explanation (logically solving Math Words Problem) – Google PaLM
  - Code generation - CoPilot
  - KG based reasoning (link prediction, message passing)
  - Image Incongruity Detection
  - Winograd Schema


Achieving AI Singularity? Not Yet
TREND5: AI Automation at scale; AI as a Service (AAAS)

- **Cloud-based pre-trained and (slightly) trainable models**
  - Easy to set up, operate and consume
  - SaaS - Mainly driven by AutoML, templates, large pre-trained models.
  - PaaS / Workbench facilities – SageMaker, Azure, Vortex, Einstein, Watson
  - Industry / Domain AI Solution (e.g., Accenture Solutions.AI)

- **Key technical issues**
  - Input Data Quality - Sample, Shape, Coverage, Bias, Class Imbalance, Anonymity and Privacy, Noise, Drift. EDA and Feature Engineering.
  - Model Quality –
    - Traditional Metrics: Precision, Recall, f-measure, AUC
    - Benchmarks
      - Single Task (BLUE, ROUGE), Multi Task (GLUE, Super GLUE), Complex Task (BIG_Bench)
  - Inference – Representation, Realism, Trust and Explainability (local – LIME, SHAP, Counterfactuals; Global – Gradient-based), Understanding Impact

Source: Datatron
Summary

In this presentation, we discussed 5 key technology innovation trends for enterprise AI. As AI has got mainstream, it is now time to assess and analyze the medium and long-term implications of these trends, especially from the point of view of standardization, interoperability, and trustworthy adoption.
Thank you

Shubhashis Sengupta

Shubhashis.Sengupta@accenture.com

https://www.accenture.com/us-en/about/accenture-labs-index

https://www.linkedin.com/in/shubhashis/
AI Quality: The Next Big Challenge in AI

Will Uppington, CEO & Co-Founder, TruEra
Introduction

Data Scientists have always focused on the performance of Artificial Intelligence/Machine Learning (AI/ML) systems. Recently, however, as the use of ML has moved from consumer to Enterprise and Government applications, increasing attention is being paid to the potential risks related to non-transparent ML systems. Enterprise organizations have struggled to move ML pilots into mainstream deployment. Regulators throughout the US and the world have issued warnings around the potential misuse of artificial intelligence in lending, employment, medical and other applications. Lawmakers, such as in the EU and New York City, have proposed or passed new laws regulating the risks of AI. To build and operate successful ML systems, developers need to account for both the performance and risks of AI/ML. This session will cover how a systematic approach to evaluating, testing and monitoring AI Quality leveraging end-to-end AI explainability technology can address the dual challenge around ML performance and risk.
Use of AI Moving from Consumer to Enterprise

**Machine Learning First Used For**

- ETA
- Search ranking
- Movie recommendation

**It’s now being used for**

- Homebuying
- Lending
- Hiring

**Classification**
- CAT

**Object Detection**
- CAT, DOG, DUCK

**Companies**
- Google
- Uber
- Netflix
- Microsoft
- Zillow
- Amazon
- Standard Chartered
- Apple
This Change Is Hurting Some Companies

Zillow to stop flipping homes for good as it stands to lose more than $550 million, will lay off a quarter of staff

Last Updated: Nov. 3, 2021 at 12:36 p.m. ET
First Published: Nov. 2, 2021 at 4:21 p.m. ET

By Jon Swartz

‘We’ve determined the unpredictability in forecasting home prices far exceeds what we anticipated and continuing to scale Zillow Offers would result in too much earnings and balance-sheet volatility,’ CEO tells investors

Hard to Build High Performing ML Models
This Change Is Hurting Some Companies

Apple Card algorithm sparks gender bias allegations against Goldman Sachs

Entrepreneur David Heinemeier Hansson says his credit limit was 20 times that of his wife, even though she has the higher credit score.
This Change Is Hurting Some Companies

Our weird behavior during the pandemic is messing with AI models

Machine-learning models trained on normal behavior are showing cracks — forcing humans to step in to set them straight.

By Will Douglas Heaven

May 11, 2020

Hard to Maintain Performance & Manage Risks Over Time
Regulators Scrutinizing AI Risks

Monetary Authority of Singapore – Fairness, Ethics, Accountability & Transparency

Hong Kong Monetary Authority – High Level Principles on AI

Banque De France – Governance of AI in Finance

National Association of Insurance Commissioners – AI guiding principles

Office of the Supdt. of FIs – Tech. Risk Consultation including AI

CFPB, OCC, Fed Reserve, FDIC, NCUS - Use of AI/ML

European Commission – Draft AI law (Credit = high risk AI)

European Insurance & Occupational Pensions Authority - AI governance

Bundesbank and BaFin – ML in Risk Models

NYC Law on Automated Employment Decision Tools


“at the CFPB, we will also be closely watching for digital redlining, disguised through so-called neutral algorithms, that may reinforce the biases that have long existed.”

- Rohit Chopra, CFPB
High Failure Rates in Enterprise AI

Failure rates for analytics, AI, and big data projects = 85% – yikes!

July 23, 2019 by Brian T. O'Neill
Enterprise AI Demands a New Approach: AI Quality

Enterprise AI requires AI Quality: balance between performance & risk mitigation
What is AI Quality?

ACCURACY

Real World Model Performance
- Conceptual Soundness
- Stability/Monitoring & Reliability
- Segment & Global Performance

Operational Factors
- Explainability & Collaboration
- Documentation

Societal Factors
- Fairness & Transparency
- Security & Privacy

Data Quality
- Missing data
- Bad data

What is AI Quality?
What is AI Quality?

- **REAL WORLD MODEL PERFORMANCE**
  - Conceptual Soundness
  - Stability/Monitoring & Reliability
  - Segment & Global Performance

- **SOCIETAL FACTORS**
  - Fairness & Transparency
  - Security & Privacy

- **OPERATIONAL FACTORS**
  - Explainability & Collaboration
  - Documentation

- **DATA QUALITY**
  - Missing data
  - Bad data
Current AI Methodologies & Tools Not Sufficient

- Limited testing, global accuracy focused; AI quality/Ethical AI not built-in
- Ad-hoc, guesswork based, lacking feedback loop from future stages
- "Black box" makes stakeholder buy-in adoption hard
- Hard to separate 'signal from noise', debug and provide feedback into next dev process

Data Prep & Engineering → Train model → Test model → Review/Buy - In/Adoption → Deploy model → Monitor model

Limited testing, global accuracy focused; AI quality/Ethical AI not built-in
Need Iterative AI Quality Tools & Mindset

Iterative AI quality review to inform data, feature & model engineering & ops buy-in

- Data Prep & Engineering
- AI Qual Review /Buy-In/Adoption
- AI Quality Evaluation
- Actionable feedback to inform iterative dev
- Ethical AI built into AI quality
- Continuous AI quality eval & debug

Train model → Deploy model → Monitor model
AI Explainability is Key to AI Quality

Black Box AI → AI Explainability → Transparent, Managed AI
AI Explainability Enables AI Quality Tooling

- Explainability
- Error Analysis
- Model Monitoring
- Bias & Fairness
- Conceptual Soundness
- Segment Analysis
- Root Cause Analysis
- Reporting
**Summary**

Succeeding in Enterprise AI requires balancing performance and risk and using AI Quality methodologies and tooling throughout the life cycle.

AI Explainability, error analysis, root cause, responsible AI and monitoring technologies can enable iterative AI quality.
Thank you

Will Uppington

will@truera.com

www.truera.com
From Data to Decisions and Back

Babak Hodjat
CTO-AI

May 24th, 2022
Another AI Winter?

An AI Timeline

- **Birth of AI**
  - Information theory
  - Digital computers
  - Cybernetics
  - The Turing Test
  - Statistical reasoning

- **Focus on Specific ‘Intelligence’**
  - Expert systems (LISP 1.5)
  - Neural networks
  - Deep learning
  - AlphaGo

- **Focus on Specific Paradigms**
  - Symbolic reasoning
  - Machine learning
  - Deep learning
  - Ontological analysis
  - Understanding language and meaning
  - Large-scale databases and algorithms
  - Experiential AI
  - High-speed computing

Timeline of AI Development

- **1950s-1960s**: First AI boom - the age of reasoning, prototype AI developed
- **1970s**: AI winter I
- **1980s-1990s**: Second AI boom: the age of Knowledge representation (appearance of expert systems capable of reproducing human decision-making)
- **1990s**: AI winter II
- **1997**: Deep Blue beats Gary Kasparov
- **2006**: University of Toronto develops Deep Learning
- **2011**: IBM’s Watson wins Jeopardy
- **2016**: Go software based on Deep Learning beats world’s champions
There is Always Room For AI

Science

- Known

Ideology

- Unknown

Software Engineering

- Programable

AI

- ML + Search + Heuristics
What Should we Expect from an AI System?
AI Should Tell Us:

- What to do
- When to do it
- What would happen if we do it
- How much should we trust it
- What would happen if we do something else…
The AI Journey Should come Full Circle

WHAT should I do?
- Optimize Outcomes with Intelligent Recommendations

WHAT will happen?
- Predict Future Events and Activities

WHY did it happen?
- Derive, Analyze and Optimize for Continuous Business Improvement

WHAT is happening?
- Develop Data Foundation and AI Preparation

- Machine Learning
- Decision-making (AI?)
- Statistical Analysis (AI?)
- Data Modernization
The Scope of AI Enablement: Step-by-step vs Wholistic

- Data
- Analytics
- Forecasts
- Predictions
- Decisions
What is wrong with this picture?

**Discovery + Planning**
- Design the solution
  1. Assess the data available within XXX
  2. Mapping of data elements
  3. Evaluate the metadata availability and fitment
  4. Finalize the technical architecture

**Data Preparation + Setup**
- Prepare the Data
  1. Setup technology Infrastructure
  2. Gather metadata and other data
  3. Engineer the metadata into required format
  4. Ingest the data and training dataset

**Asset Evaluation**
- MVP for aiding asset evaluation
  - For a job input by the agency, predict:
    1. How likely it is to pass each dimension of evaluation along with
    2. An estimation of certainty in this prediction
    3. Identify and mark the probable areas of concern on the asset

**Campaign Effectiveness**
- MVP for aiding campaign effectiveness
  - Predict whether:
    1. The job is likely to receive regulatory complaints after it has been approved
    2. The campaign will be effective
    3. Identify and mark areas of probable concern on the asset

**Suggestion**
- MVP for aiding asset design
  - Suggest how:
    1. Elements of the job could be improved
    2. Allow users to edit the suggestions
    3. Predict how likely is it for the edits to pass evaluations
      a. Receive complaints, and
      b. Perform well in the marketplace
AI Hesitancy

Typical excuses for deferring or rejecting AI-enablement
Our data is a mess/not ready

<table>
<thead>
<tr>
<th>Timestamp</th>
<th>Rank 1</th>
<th>Rank 2</th>
<th>Rank 3</th>
<th>Rank 4</th>
<th>Rank 5</th>
<th>Rank 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>12/5/2016 13:20:15</td>
<td>22 A Million, Bon Iver</td>
<td>Coloring Book, Chance the R We've Got It From Here, Thx Moon, I'm a Pop Lover, The Avett Brothers, Anythings, Jame</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12/5/2016 13:20:19</td>
<td>The Posies, Solid States</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12/5/2016 13:20:28</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12/5/2016 13:20:31</td>
<td>ZAYN, Mind of Mine</td>
<td>Ariana Grande, Dangerous W, Solange, A Seat at the Table</td>
<td>Chance the Rapper, Coloring dsvn, SEPT. 5TH</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12/5/2016 13:21:01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12/5/2016 13:21:11</td>
<td>Hope Sandoval, Until the Hur Madness, Can't Touch Us No</td>
<td>Suzanne Vega, Lover, Beloved Poliça, United Crushers</td>
<td>Underworld, Barbara Barbara</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12/5/2016 13:21:40</td>
<td>Beyonce, Lemonade</td>
<td>Radiohead, A Moon Shaped</td>
<td>Parquet Courts, Human Perfc Cate Le Bon, Crab Day</td>
<td>Nels Cline, Lovers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12/5/2016 13:21:59</td>
<td>Josh Ritter, Sermon on the R. Lor McKeeman, The B&amp;G, Sara Watkins, Young In All Ways</td>
<td>Sara Watkins, Young In All Ways</td>
<td>Beyonce, Formation, Anti, Jojo, Mad Love</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12/5/2016 13:23:06</td>
<td>Tancred, Out of the Garden</td>
<td>Jimi Eat World, Integrity BL, Sam Beam &amp; Jess, Hoodlum, Venerable, Young</td>
<td>Sara Watkins, Young In All Ways</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12/5/2016 13:25:10</td>
<td>RVRB, FADED</td>
<td>RVRB, FADED</td>
<td>RVRB, FADED</td>
<td>RVRB, FADED</td>
<td>RVRB, FADED</td>
<td>RVRB, FADED</td>
</tr>
<tr>
<td>12/5/2016 13:25:26</td>
<td>Care Le Bon, Crab Day</td>
<td>David Bowie, Blackstar</td>
<td>Weaves, Weaves</td>
<td>Trentmoller, Fixion</td>
<td>Solange, a Seat at the Table</td>
<td></td>
</tr>
<tr>
<td>12/5/2016 13:25:28</td>
<td>Cars Headrest, Teens of E Beyonce, Lemonade</td>
<td>Radiohead, Moon Shaped Pc Margaret Glaspy, Emotions a Dawes, We're All Gonna Die</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
We don’t have enough Data
More data does not mean, more features!
The Curse of Dimensionality

More features means exponentially more observations to avoid overfitting
Outcomes Timeseries Captures Effect of Missing Features

AI’s Predicted New Cases per Day in United States / California

7-day moving average forecast w/ current NPIs held out
Confidence
7-day moving average forecast w/ AI’s suggested NPIs
Historical 7-day moving average
Confirmed New Cases

Date

New Cases
0 10k 20k 30k 40k 50k
No One-Size-Fits-All AI

Domain Specific Models are the new IP
Need to get Insights First
What if Insights Show we are Good Already?
Human Decision-Making is Better
All we need is better insights - we can make the decisions

Projections Show we’re Missing Revenue Targets!

Should we cut costs?
Sell at a discount?
Retarget sales?
Invest in growth?
We care about sales – Cost is for a different department

There is never a single objective!

We often want to:

• Maximize revenue, and
• Maximize sales, while
• Minimizing cost, and,
• Minimizing risk
Worried about bias and ethics of using AI

“And finally this is Phil Goodwin. He’s part of the problem.”
We make Decisions all the time

- Given a context (C)
- Which actions should I take (A)
- To optimize some outcomes (O)?
Adaptive Decision-Making: The OODA Loop
How do we make decisions?

• Given a **context** (C)
• Which **actions** should I take (A)
• To optimize some **outcomes** (O)?
AI-Enabled Decisioning (and Data Collection)

- Observe
- Model
- Suggest
- Implement
Important Questions AI can Answer

**Predictor:**
What will happen?

**Prescriptor:**
What should we do?

**Uncertainty:**
How much can we trust it?

**Scratchpad:**
What-if we do something else?

**Impact Estimation (Retrospective):**
What if we had done something else?

**Multiple-Objectives:**
What if we want a different balance of outcomes?
How it all comes together

Data collects

Data aggregation

Historical Decision Data

CAO

Data Science

Predictor

Surrogate Model / Simulator

C & A \(\rightarrow\) O

Prescriptor

C \(\rightarrow\) A

Certainty Model

Trust

Certainty Model

Decision-maker applies

A

O collects data

Decision-maker applies certainty model
Digital Twins

Physical environment

Digital Twin Instance

Prediction and Optimization

- Simulation
- What-if scenarios
- Prediction and prescription
- Optimization
- Predictive maintenance
- Remaining useful life

Design Systems
- Enterprise Data
- Sensor Data
- Historical data
- Image data
- Acoustic data

Vibration
- Heating Temperature
- Flow Speed
- Viscosity
- Vapor pressure
- Power Consumption monitor
- Chamber humidity
- Current and voltage

………..
Rigid Simulations

Make Assumptions
- May be unavailable
- May be wrong
- May be inaccurate
- May be insufficient

Create Model
- Requires knowledge of mechanisms and dynamics

Run Model
- Often expensive at scale
- Often inaccurate in miniature

Interrogate Model
- Rigid

Digital Twin

Concept diagram of climate modeling

SEIR: Susceptible Exposed Infected Recovered Compartment Model

Source: 2000 W.F. Ruddiman
ML Models

Robust
Data Driven
Does not require (as much) knowledge of the domain
Can be applied at scale

BUT not enough
Data-hungry
Training could be expensive/time consuming
Not adaptive
Gradient Descent is not good at large, deceptive, multi-objective search spaces
Why use Evolutionary AI?

- A population of encodings (e.g., lists or trees)
- Decoded into individuals that are tested
- Good individuals retained, bad thrown away

- New individuals generated from the parent encodings
  - Crossover: Combine building blocks from two parents
  - Mutation: Create new building blocks
Evolution is Efficient

- Works in large spaces (e.g., $2^{2^{70}}$)
- Multiple variables optimized at once (e.g., up to 1B)
- Multiple objectives and novelty get around deception
Evolution can handle Multiple Objectives

- Animation: Pareto front by generation for single-objective (green) vs. multi-objective (blue)
- Single-objective focuses on improving largest networks
- Multi-objective focuses on improving the entire curve
- Result: Multi-objective finds much smaller models for the majority of performance values.
Evolution Can Create Explainable Solutions

- 3_53 Rules agent:
  - `<22978> (0.11*S_1^3 < 0.87*S_2) ---> LEFT`
  - `<23022> Default Action: RIGHT`
Each rule is a variable length conjunction of conditions with an associated class prediction. In the evaluation each condition compares a lagged value or the current value of the time series to a threshold (decision boundary). Decision lists in LEAF have a variable number of rules and conjunctive clauses in each rule.

This representation is different from many other classifiers e.g. Decision Trees, simple Decision Lists, Support Vector Machines and Logistic Regression, which requires every time lagged value or an aggregate to be explicitly presented as a different feature.
Diabetes Rules

1. \( (\text{diag}_1\_\text{injury} \leq 2.2\times \text{diag}_1\_\text{respiratory}) \) and \( (\text{diag}_2\_\text{neoplasms} > 33) \) —> 18

2. \( (\text{age}_{[60-70]} \geq 1.2\times \text{age}_{[10-20]}) \) —> 7

3. \( (\text{diag}_2\_\text{diabetes} \geq 0.82\times \text{diag}_1\_\text{digestive}) \) —> 17

4. \( (\text{admission\_source\_id\_Court/Law Enforcement} \geq 1.045\times \text{diag}_3\_\text{circulatory}) \) —> 9

5. \( (\text{age}_{[80-90]} \geq 2.29\times \text{admission\_type\_id\_Newborn}) \) —> 14

6. \( (\text{age}_{[30-40]} \leq 0.47\times \text{age}_{[50-60]}) \) and
   \( (\text{admission\_type\_id\_Newborn} \leq 0.47\times \text{age}_{[50-60]}) \) and
   \( (\text{diag}_2\_\text{respiratory} \geq 1.86\times \text{diag}_3\_\text{musculoskeletal}) \) and
   \( (\text{diag}_1\_\text{diabetes} \geq 1.42\times \text{age}_{[30-40]}) \) and
   \( (0.18\times \text{gender} \geq 0.38\times \text{race\_Asian}) \) and
   \( (0.12\times \text{race\_AfricanAmerican} \geq 0.15\times \text{admission\_source\_id\_Court/Law Enforcement}) \) and
   \( (0.03\times \text{race\_Hispanic} \leq 0.76\times \text{admission\_source\_id\_Emergency Room}) \) —> 15

7. Default 8
Diabetes Rules Interpretation

1. If Patient has a respiratory problem but not a neoplasms problem prescribe metformin-pioglitazone

2. If the person any age except 60-70 prescribe glyburide

3. If the person has Diabetic Diagnostic result as positive prescribe glipizide-metformin

4. If the person has a positive circulatory diagnosis prescribe pioglitazone

5. If the admission type is Newborn prescribe tolazamide

6. If the person is between 30 and 60 years old and has a positive musculoskeletal diagnosis and is Asian and was sent by court/law enforcement and was admitted in emergency, prescribe insulin

7. If none of the above, prescribe glipizide
Some Examples
Pricing Optimization

Improving pricing strategy for a leading home appliance manufacturer

- Using historical data, build an adaptive digital twin of the market
- With evolutionary optimization, maximize conflicting objectives of revenue and margin

Result: A Pareto front of the best possible tradeoffs
- The user can select a desired tradeoff from it
- RIO indicates uncertainty in the predictions

Scratchpad allows adjusting the price by hand
- The digital twin predicts the resulting revenue and margin
Insurance Underwriting

Improving pricing strategy for a leading property insurance company

- Using historical data, build an adaptive digital twin of the properties
- With evolutionary optimization, maximize conflicting objectives of win/loss, risk, and cost

Result: A Pareto front of the best possible tradeoffs
- The user can select a desired tradeoff from it
- RIO indicates uncertainty in the predictions

Scratchpad allows adjusting the premiums by hand
- The digital twin predicts the resulting risk and win/loss
Ship Trim Optimization

Improving ship speed/fuel efficiency for a shipping company

- Using historical data, build an adaptive digital twin of the ship
- With evolutionary optimization, maximize conflicting objectives of speed and fuel consumption

Result: A Pareto front of the best possible tradeoffs
- The user can select a desired tradeoff from it
- RIO indicates uncertainty in the predictions

Scratchpad allows adjusting the premiums by hand
- The digital twin predicts the resulting speed and fuel consumption
Optimizing Growth Recipes for Agriculture

- Build an adaptive digital twin of how plants grow in hydroponic growth environments
  - Light, temperature, water, nutrients etc. computer controlled
- Evolve growth recipes against the twin
  - Millions new recipes created and evaluated; only a few hundred actually planted
- Discovered that basil does not need to sleep!
  - Grows best when lights are on 24hrs/day
  - An insight that surprised domain experts
Optimizing COVID-19 Response

Build an adaptive digital twin of COVID-19 cases over time
- Predict number of cases in different countries
- In response to non-pharmaceutical interventions (NPIs)

Discover best interventions against the twin
- Resulting in smallest number of cases
- With minimal economic cost

Not just what will happen, but what we should do about it!
- Scratchpad, uncertainty to help explore alternatives

Retrained continuously since May 2020
- Adapting to the different stages of the pandemic
- Generalizing from experiences across the world

Recommendations about two weeks in advance, e.g.
- May 2020: Focus on schools and workplaces (i.e. indoors)
- Sept 2020: Focus on gatherings, travel restrictions
- March 2021: India lockdown
- July 2021: Delta surge on countries with low rates so far
- March 2022: Masking to avoid a second Omicron surge

Interactive demo: https://evolution.ml/demos/npidashboard
An Organization Consists of Many Decision Makers
Scale in Entangled Decision-Loops

- More Accurate Decision
- More Efficient Decision-making
- Better Outcomes
Thank You
How specialized AI drives value for Automotive Service Organizations

Tilak Kasturi
Automotive Ecosystem Use Cases

- Fleets
- Insurers
- OEMs
- Dealers
- Aftermarket Service Providers
- Government Agencies
- Parts Suppliers
- Consumers

**Manufacture**
- Vehicle Sales Forecast
- Vehicle Design
- New / Used Vehicle Sale Prices
- Production Costs
- OEM Repair Procedures
- Parts Pricing (Listed / Paid)

**Sell**
- Residual Value
- Parts Production Forecast
- Labor Rates Incurred
- Replacement Part SKUs
- Vehicle Design Specifications
- Repair Cycle Time

**Operate**
- Repair Diagnostics
- Repair Estimation
- Parts Replacement Occurrence
- Driver Behavior
- Collision Severity
- Parts Performance and Wear

**Repair**
- Predictive Maintenance

**Remarket**
- Usage Based Insurance

**Salvage**
- Road Capacity Planning
- Diagnostic Trouble Codes
- Emissions Levels
- Shop Equipment Utilized

Source: Jefferies
Automotive Data is unstructured, complex.

- Trillion total global market
- $2 trillion
- 2% of US nominal GDP
- 4.7 million jobs in the US
- Billion vehicles in operation globally
- ~1.4 billion
- 12.2 years average age of light vehicles (US cars & light trucks)
- 100K make, model, engine, transmission combinations
- 2% of US nominal GDP
- 4.7 million jobs in the US
Automotive service focus is to minimize breakdowns

Predict and Prescribe

Right Part, right place, right time
Challenges with Automotive service data for AI Applications

• Availability of data for research community
• Proprietary and heterogeneous data
• Large NLP models are not trained on “automotive data”
• Data is too noisy
  • Textual data has industry and local acronyms, incomplete sentence structures, etc.
  • Connected car data is not continuous, lacks coverage
State-of-the-art NLP Trends

• Transformers changed the landscape and is here to stay
• Thrust on low-resource languages; multilingual NLP going mainstream
• Breaking barriers: Supervised --> Unsupervised --> Self-supervised learning (hybridization)
• Low-code NLP on the rise
• Multimodal systems gaining momentum
Advancements in developing large NLP Models

- BERT started the trend towards dynamic models
- Race towards larger NLP models
- Megatron-Turing NLG model at 530B parameters!
- Powering downstream tasks
  - Semantic search
  - Automated dialogue generation
  - Summarization
  - Machine translation
  - Commonsense reasoning

Figure 1. Trend of sizes of state-of-the-art NLP models over time

Source: Using DeepSpeed and Megatron to Train Megatron-Turing NLG 530B, the World’s Largest and Most Powerful Generative Language Model - Microsoft Research
Large Language Models are not meant for industry specific use-cases

- Models should drive specific decisions
- Models need to be explainable
- Models should have high confidence
- Models need ‘deep contextual understanding’ in specialized domain

*When will breakdowns happen and why?*
*Which repair will resolve the issue?*
*Which issues are a safety concern?*
MLOps & Cloud native AI Pipeline drive actionable decisions

Data source
- Data normalization
  - Diagnostic code discovery
  - Feature Discovery
  - Labor operations and component discovery
  - Domain Knowledge Filtering
  - Classification and Clustering
  - Results UI/API

16272: P1110: Labor - Noise in engine
- Removed oil pan and balance shafts to inspect crank. Found high amount of meta debris in the pan. Found the #2 rod bearing had spun and chewed up the crankshaft and rod.

18213: Labor - LONG BLOCK - Remove & Replace - LX - Consists of a cylinder block fitted with Pistons, Rings, Connecting Rods, Crankshaft, and all Bearings, Cylinder Head(s), Camshaft(s), Timing Chain or Belt and Sprocket(s) or Gears. <inlude where applicable>: Clean

RO Lines
18211: Part - Oil Filter
18219: Part - Oil
18220: Part - Oil Filter

Failure, Engine Oil Filter Failed

MAHCO Model: M5 2011 Year: 2012 Mileage: 11,048 Reported Date: 2021-04-08 Complaint Source: Vehicle Owners' Questionnaire [Website]

MAHCO Model: M5 2011 Year: 2012 Mileage: 11,048 Reported Date: 2021-04-08 Complaint Source: Vehicle Owners' Questionnaire [Website]

MAHCO Model: M5 2011 Year: 2012 Mileage: 11,048 Reported Date: 2021-04-08 Complaint Source: Vehicle Owners' Questionnaire [Website]

MAHCO Model: M5 2011 Year: 2012 Mileage: 11,048 Reported Date: 2021-04-08 Complaint Source: Vehicle Owners' Questionnaire [Website]

Failure, Throttle Body Failed

MAHCO Model: TS722 2011 Year: 2012 Mileage: 11,048 Reported Date: 2021-04-08 Complaint Source: Vehicle Owners' Questionnaire [Website]

Predict Confidential
Two examples of Future trends in Automotive Service

“John, Your 2019 BMW X3 needs service”

“What are my vehicle’s top Safety Concerns?”
“John, Your 2019 BMW X3 needs service.”

1. Vehicle IOT Trigger
   - DTC, PID via Telematics

2. Upstream Analytics
   - Min, Max, Anomalies isolated

3. Overlay & Correlate
   - Overlaying with insights from repair orders (textual)
   - DTC(P0171,P0440), Short Fuel Trims(%), O2 Sensor(v)

4. Predict Parts Replacement
   - Vehicle YMME & mileage specific

5. Pro-active CRM

6. Parts Ordering
   - BMW Vapor Canister Purge Solenoid - Genuine BMW 13907618647
   - $97.99

7. Vehicle Repair
“What are my vehicle’s top Safety Concerns?”
Important Conferences and Journals

- ACL
- EMNLP
- COLING
- NAACL
- EACL
- AACL-IJCNLP
- TACL
- CL
- AAAI, IJCAI, NeurIPS, CVPR, ICML, ICLR, ICCV, ECCV
- + many relevant workshops

Twitter: #NLProc
Thank you

Tilak Kasturi
tilak@predii.com
www.predii.com