Guiding AI to Benefit Humanity and the Environment

GONG Ke

World Federation of Engineering Organizations
Over my lifetime, I have seen very significant societal changes. Probably one of the most significant, and one that is increasingly concerning people today, is the rise of artificial intelligence. In short, I believe that the rise of powerful AI, will be either the best thing, or the worst, ever to happen to humanity. I have to say now, that we do not yet know which. But we should do all we can, to ensure that its future development benefits us, and our environment. We have no other option.

Guiding AI to Benefit humanity and the environment

Engineering for Sustainable Development
1956, “artificial intelligence” had been proposed for “analogy of human intelligence by machine”

2006, AI@50, AI was still not useful
Breakthrough: Deep Learning + Big Data

SUN, 131K
[Xiao et al. '10]

LabelMe, 37K
[Russell et al. '07]

PASCAL VOC, 30K
[Everingham et al. '06-'12]

Caltech101, 9K
[Pei-Fei, Fergus, Perona, '03]

Engineering for Sustainable Development
Facial Recognition

Engineering for Sustainable Development
Predictive policing, real-time crime mapping, and gunshot detection have the greatest impact on preventing deaths. In a city with the population and crime profile of Rio, this could mean saving some 300 lives each year. Incidents of assault, robbery, and burglary could be lowered by 30–40 percent.
In 1950, A. Turing had devised a Turing Test in his paper of “Computation Machine and Intelligence”, and predicted that it would be passed by 2000. However, until 7th of June 2014, the 60 years memorial day of the death of Alan Turing, the Chatbot Eugene Goostman has passed Turing Test for the first time.
I'm already very interested in design, technology and the environment.

Engineering for Sustainable Development
Engineering for Sustainable Development
Speech recognition and translation

Engineering for Sustainable Development
15-30 min shaved off average daily commute

Hangzhou in Traffic-clotted cities
2015: 4th in China
2018: 57th in China
Average commuter time reduced 15%, emergence traffic speed doubled
On July 10, Tsinghua University announced a survey to young people in China, it shown that >70% of the surveyed people have used more than 3 kinds of AI products. This tells that AI is going out of labs and papers but into the scenes of our everyday life and work.
Major challenges to AI development

Data Dependency
Non-transferability
Low Energy efficiency
Semantic Gaps
Not explainable
Reliable?
Risks and challenges

• Artificial intelligence has a wide range of far-reaching influences and has the dual attributes of technology and society. It may bring about changes in the
  • structure of employment,
  • impact on legal and social ethics,
  • infringe on individual privacy, and
  • challenge the norms governing international relations.

It will have far-reaching implications for governance, economic security and social stability and even global governance.
Discriminative behavior of AI

The criminal risk assessment algorithm used by Chicago court has discrimination

- COMPAS: Correctional Offender Management Profiling for Alternative Sanctions algorithm

<table>
<thead>
<tr>
<th></th>
<th>HUMANS</th>
<th>COMPAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy (overall)</td>
<td>67.0%</td>
<td>65.2%</td>
</tr>
<tr>
<td>False positive (black defendants)</td>
<td>37.1%</td>
<td>40.4%</td>
</tr>
<tr>
<td>False positive (white defendants)</td>
<td>27.2%</td>
<td>25.4%</td>
</tr>
<tr>
<td>False negative (black defendants)</td>
<td>29.2%</td>
<td>30.9%</td>
</tr>
<tr>
<td>False negative (white defendants)</td>
<td>40.3%</td>
<td>47.9%</td>
</tr>
</tbody>
</table>

J. DRESSEL ET AL., SCIENCE ADVANCES, EAAO5580, 2018, ADAPTED BY C. AYCOCK/SCIENCE

*Engineering for Sustainable Development*
Adversarial test

Engineering for Sustainable Development
Proposed Principles for Responsible Conduct with Big Data and AI in Engineering

1. Good for Humanity and Its Environment
2. Fairness, Inclusiveness and Public Awareness
3. Privacy and Data Integrity
4. Transparency
5. Accountability
6. Peace, Safety and Security
7. Collaboration
## AI governance techniques

<table>
<thead>
<tr>
<th>Ethical Requirement</th>
<th>AI Governance Techniques</th>
</tr>
</thead>
<tbody>
<tr>
<td>Security and Privacy</td>
<td><em>Federated Learning, Blockchain</em></td>
</tr>
<tr>
<td>Safety and Reliability</td>
<td><em>Machine Learning Test and Verification</em></td>
</tr>
<tr>
<td>Transparency</td>
<td><em>Interpretable / Explainable AI</em></td>
</tr>
<tr>
<td>Accountability</td>
<td><em>AI Provenance, Auditing and Forensic</em></td>
</tr>
<tr>
<td>Fairness</td>
<td><em>AI Fairness Evaluation and Debiasing algorithm</em></td>
</tr>
</tbody>
</table>
Test, Verification, Interpretation and Provenance for Trustworthy AI

Engineering for Sustainable Development
AI DevOps for Complete Governance

deploy the model in the system, examine whether the ethical constraints are violated in operation, and make decisions on further model improvement.

Through requirement analysis to choose the right ethical specification and generate the AI task definition.

Ensure the validity of the training dataset, reduce the potential bias and achieve data balance.

According to the ethical specification, design model structure and perform model training.

check whether the trained model satisfy the ethical specification (fairness, robustness, transparency and quality).

Test and Verify Model

Design and Train Model

Deploy and Apply Model

Data Collection and Preparation

AI Task Definition

Engineering for Sustainable Development
AI with human values for sustainable development

Artificial intelligence with human values for sustainable development

“Artificial intelligence can be a great opportunity to accelerate the achievement of sustainable development goals. But any technological revolution leads to new imbalances that we must anticipate.”