Regulating Artificial Intelligence? Reflecting on Several International Recommendations

James Larus

School of Computer and Communications Sciences
EPFL

Vienna Workshop on Digital Humanism
April 4-5, 2019
Faculty of Informatics - TU Wien
Vienna, Austria
Question for the audience

• In 2019, should your government regulate the use of Artificial Intelligence (AI) in commercial products?
Is your view universal? common?
COMPUTING MACHINERY AND INTELLIGENCE

By A. M. Turing

I. The Imitation Game

I propose to consider the question, "Can machines think?" This should begin with definitions of the meaning of the terms "machine" and "think." The definitions might be framed so as to reflect as far as possible the normal use of the words, but this attitude is dangerous, if the meaning of the words "machine" and "think" are to be found by examining how they are commonly used it is difficult to escape the conclusion that the meaning and the answer to the question, "Can machines think?" is to be sought in a statistical survey such as a Gallup poll. But this is absurd. Instead of attempting such a definition I shall replace the question by another, which is closely related to it and is expressed in relatively unambiguous words.

The new form of the problem can be described in terms of a game which we call the "imitation game." It is played with three people, a man (A), a woman (B), and an interrogator (C) who may be of either sex. The interrogator stays in a room apart from the other two. The object of the game for the interrogator is to determine which of the other two is the man and which is the woman. He knows them by labels X and Y, and at the end of the game he says either "X is A and Y is B" or "X is B and Y is A." The interrogator is allowed to put questions to A and B thus:

C: Will X please tell me the length of his or her hair?

Now suppose X is actually A, then A must answer. It is A's object in the game to try and cause C to make the wrong identification. His answer might therefore be:

"My hair is shingled, and the longest strands are about nine inches long."
Applied AI

Specific (human) skills
ImageNet is an annual contest to automatically classify images in 1000 categories.

Error rate for human is ~5% (but people can handle many more categories).

Whole is more than the sum of its parts – Aristotle (Metaphysics, Book 8)
Why regulate AI?

- AI is dangerous
- AI differs from other technologies
- Cannot predict technical/social consequences
- Will continue/exacerbate social inequities
- Singularity/Superintelligence
- Enhance public acceptance
Why not regulate AI?

• Dangers overestimated
• Can manage consequences
• Lack sufficient understanding
• Impossible to regulate
• Prevent beneficial uses
• Lose commercial competition
Three caricatures

Europe – Moralist

Anglo-Saxon – Capitalist / legalist

Chinese – Entrepreneurial / authoritarian

In interest of exposition, I will ignore subtlety and divergence of opinions. 😊
Europe
European Commission High-Level Expert Group on Artificial Intelligence (AI HLEG)

- “AI’s benefits outweigh its risks
- “Follow the road that maximizes the benefits of AI while minimizing its risks
- “Human-centric approach to AI is needed
  - Goal to increase human well-being
- “Three pillars underpin the Commission’s vision: (i) increasing public and private investments in AI to boost its uptake, (ii) preparing for socio-economic changes, and (iii) ensuring an appropriate ethical and legal framework to strengthen European values.
Trustworthy AI

• “AI is thus not an end in itself, but rather a means to increase individual and societal well-being.

• “Trust is a prerequisite for people and societies to develop, deploy and use Artificial Intelligence.”
Trustworthy AI will be our north star

- “It should respect fundamental rights, applicable regulation and core principles and values, ensuring an ‘ethical purpose’

- “It should be technically robust and reliable since, even with good intentions, a lack of technological mastery can cause unintentional harm

- “Guidelines are not meant to stifle AI innovation in Europe, but instead aim to use ethics as inspiration to develop a unique brand of AI, one that aims at protecting and benefiting both individuals and the common good.

- “This allows Europe to position itself as a leader in cutting-edge, secure and ethical AI."
“Use ethics to inspire trustworthy development, deployment and use of AI.

Use the fundamental rights commitment of the EU Treaties and Charter of Fundamental Rights as the stepping stone to identify abstract ethical principles, and to specify how concrete ethical values can be operationalized in the context of AI.

i.e., dignity, freedoms, equality and solidarity, citizens’ rights and justice

“Respect for fundamental rights, principles and values – and ensuring that AI systems comply therewith – is coined here as ensuring ‘ethical purpose’
Principle of beneficence: “do good”

- “AI systems can contribute to well being by seeking achievement of a fair, inclusive and peaceful society, by helping to increase citizen’s mental autonomy, with equal distribution of economic, social and political opportunity.

- “AI systems can be a force for collective good when deployed towards objectives like:
  - the protection of democratic process and rule of law;
  - the provision of common goods and services at low cost and high quality;
  - data literacy and representativeness;
  - damage mitigation and trust optimization towards users;
  - achievement of the UN Sustainable Development Goals or sustainability understood more broadly, according to the pillars of economic development, social equity, and environmental protection.
Requirements for trustworthy AI

- Accountability
- Data Governance
- Design for all
- Governance of AI Autonomy (Human oversight)
- Non-Discrimination
- Respect for Human Autonomy
- Respect for Privacy
- Robustness
- Safety
- Transparency
• “The primary purpose of partly and fully automated transport systems is to improve safety for all road users

• “The licensing of automated systems is not justifiable unless it promises to produce at least a diminution in harm compared with human driving

• “technologically unavoidable residual risks do not militate against the introduction of automated driving if the balance of risks is fundamentally positive.

• “Automated and connected technology should prevent accidents wherever this is practically possible

• “Liability for damage caused by activated automated driving systems is governed by the same principles as in other product liability.
Europe

- Thoughtful, well intentioned reports

- Ethical approach: AI should be better than the most upright, moral, humane person

- Devolves responsibility onto implementers (engineers and business people)
  - Not a call for new regulation
  - Standardization, accountability governance, codes of conduct, education, dialogs, diversity of teams

- Not much consideration of social impact

- Goals, not rules, for engineering AI
Point/Counterpoint

**Point:** Should AI Technology Be Regulated? Yes, and Here’s How.

Considering the difficult technical and social questions affecting the regulation of artificial intelligence research and applications.

**Counterpoint:** Regulators Should Allow the Greatest Space for AI Innovation

Permissionless innovation should be the governing policy for AI technologies. It is counterintuitive, but this is the well-established calculus of the world of risk analysis. While policymakers are risk-averse in general and may have good reason to be so, it is also possible to go too far. For example, policymakers often equate risk with uncertainty, but there are many situations in which risk and uncertainty are actually inversely proportional. As the number of studies grows, so does our knowledge of the potential harms of AI systems, it is important that public policy not freeze the development of life-enriching innovations in this space based on speculative fears of an uncertain future. When we consider the future of AI technologies, it is vital to keep that in sight in mind. AI-enabled technologies can pose some risks that should be taken seriously, it is important that policymakers carefully consider the potential benefits AI technologies can bring in vehicle safety, improved productivity, and much more.

**Oren Etzioni**

**Counterpoint:** Regulators Should Allow the Greatest Space for AI Innovation

Permissionless innovation should be the governing policy for AI technologies. It is counterintuitive, but this is the well-established calculus of the world of risk analysis. While policymakers are risk-averse in general and may have good reason to be so, it is also possible to go too far. For example, policymakers often equate risk with uncertainty, but there are many situations in which risk and uncertainty are actually inversely proportional. As the number of studies grows, so does our knowledge of the potential harms of AI systems, it is important that public policy not freeze the development of life-enriching innovations in this space based on speculative fears of an uncertain future. When we consider the future of AI technologies, it is vital to keep that in sight in mind. AI-enabled technologies can pose some risks that should be taken seriously, it is important that policymakers carefully consider the potential benefits AI technologies can bring in vehicle safety, improved productivity, and much more.

**Permissionless innovation should be the governing policy for AI technologies.**

James Larus
Americans are concerned about AI

An overwhelming majority of Americans think that AI and robots should be carefully managed.

<table>
<thead>
<tr>
<th>Opinion</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Totally agree</td>
<td>52%</td>
</tr>
<tr>
<td>Tend to agree</td>
<td>30%</td>
</tr>
<tr>
<td>Tend to disagree</td>
<td>8%</td>
</tr>
<tr>
<td>Totally disagree</td>
<td>1%</td>
</tr>
<tr>
<td>I don’t know</td>
<td>12%</td>
</tr>
</tbody>
</table>

Americans trust tech companies and non-government organizations more than the government to manage AI.

<table>
<thead>
<tr>
<th>Organization</th>
<th>Great deal of confidence</th>
<th>Fair amount of confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tech companies</td>
<td>8%</td>
<td>33%</td>
</tr>
<tr>
<td>Partnership on AI</td>
<td>9%</td>
<td>30%</td>
</tr>
<tr>
<td>Non-government scientific organization (e.g. AAAI)</td>
<td>8%</td>
<td>30%</td>
</tr>
<tr>
<td>Intergovernmental research organizations (e.g. CERN)</td>
<td>7%</td>
<td>31%</td>
</tr>
<tr>
<td>International organizations</td>
<td>6%</td>
<td>22%</td>
</tr>
<tr>
<td>UN</td>
<td>6%</td>
<td>22%</td>
</tr>
<tr>
<td>US federal government</td>
<td>7%</td>
<td>20%</td>
</tr>
<tr>
<td>US state governments</td>
<td>6%</td>
<td>20%</td>
</tr>
</tbody>
</table>
Oren Etzioni

- CEO Allen Institute for AI and Prof. University of Washington

“Government regulation is necessary to prevent harm. But regulation is also a blunt and slow-moving instrument that is easily subject to political interference and distortion. When applied to fast-moving fields like AI, misplaced regulations have the potential to stifle innovation and derail the enormous potential benefits that AI can bring in vehicle safety, improved productivity, and much more. We certainly do not want rules hastily cobbled as a knee-jerk response to a popular outcry against AI stoked by alarmists such as Elon Musk...
“avoid regulating AI research, but move to regulate AI applications in arenas such as transportation, medicine, politics, and entertainment.

“The AI train has left the station; AI research will continue unabated and the U.S. must keep up with other nations or suffer economically and security-wise as a result.
Guidelines for regulators

1. “Don’t weaponize AI
2. AI is subject to full gamut of laws that apply to its human operator.
3. AI should clearly disclose that it is not human.
4. AI shall not retain or disclose confidential information without explicit prior approval from the source.
5. AI must not increase any bias that already exists in our systems.

• “Regulatory bodies will enact ordinances, or legislators will enact laws to protect us from the negative impacts of AI in applications.
Paradoxically, sometimes the policies we implement to guarantee our safety end up making us much worse off than if we had done nothing at all. It is counterintuitive, but this is the well-established calculus of the world of risk analysis.

It is important that public policy not freeze the development of life-enriching innovations in this space based on speculative fears of an uncertain future.

When considering policy for AI and related emerging technologies such as robotics and big data, policymakers face two general options regarding how best to respond to new technological developments: They can either choose to preemptively set limits or bans on new technologies if they believe the risks to society are simply too great to tolerate—an approach known as the “precautionary principle”—or they can decide to allow innovation to proceed mostly unhampered and intervene only in a post hoc or restitutionary manner, which we call “permissionless innovation.”

We believe artificial intelligence technologies should largely be governed by a policy regime of permissionless innovation so that humanity can best extract all of the opportunities and benefits they promise. A precautionary approach could, alternatively, rob us of these life-saving benefits and leave us all much worse off.

The Risk of Avoiding All Risks

Human psychology is such that the precautionary principle often initially seems appealing. We, as a species,
Recommendations

• “Articulate and defend permission-less innovation as the general policy default.
• Identify and remove barriers to entry and innovation.
• Protect freedom of speech and expression.
• Retain and expand immunities for intermediaries from liability associated with third-party uses.
• Rely on existing legal solutions and the common law to solve problems.
• Wait for insurance markets and competitive responses to develop.
• Push for industry self-regulation and best practices.
• Promote education and empowerment solutions and be patient as social norms evolve to solve challenges.
• Adopt targeted, limited, legal measures for truly hard problems.
• Evaluate and reevaluate policy decisions to ensure they pass a strict benefit-cost analysis.
Regulation in (Anglo-Saxon) legal system

Chris Reed, Professor, School of Law, Queen Mary University of London

“Good regulation would improve our perception of safety, and also our perception that humans remain in control. It could also mitigate any new risks which the use of AI creates.

“And the ways in which AI gets it wrong are likely to be very different from the ways in which a human would make mistakes. This feels dangerous to society. We want to know the kinds of risks we are running, and purely statistical arguments that AI makes us safer are not convincing to the wider population.

“But bad regulation risks stifling the development and implementation of useful AI solutions, perhaps even without improving safety and control.

Liability law

• Outside specifically regulated sectors, the general approach of law and regulation is that innovation is freely permitted, but that those responsible must bear the consequences if that innovation causes certain types of harm.
Against new laws

- “Regulation cannot control unknown risks, and devising a regulatory mandate on the basis of speculative risks seems unlikely to produce successful results.

- “lawmakers are generally unsuccessful at prospective regulation, particularly in technology fields.

- “a regulatory regime which aimed to deal with all uses of AI technology would be impossibly wide in scope.
Fundamental rights

• “Most countries have a range of laws which protect fundamental rights. Those most obviously threatened by AI decisions are the rights not to be discriminated against on the grounds of race, sex, sexual orientation, religion, etc.

• “The current focus of the law places the obligation to avoid human rights infringement on the person who is ultimately responsible for the decision, and not on the technology producer.
Allocation of responsibility

• “In a minority of cases, the person responsible for the activity is made **strictly liable**, which means that they must compensate those who suffer loss or damage, regardless of whether the person responsible was careless when undertaking the activity.

• “In most cases, the applicable responsibility regime will be the **law of negligence**, which is explicitly based on fault. This imposes liability for loss or damage caused by carelessness.
Autonomous vehicles

• “If we take the example of autonomous vehicles, it is immediately apparent that the nature of negligent driving changes. We can no longer ask whether the person driving was careless, because no person is driving.

• “How was the accident caused? and, Was the person who was responsible for the cause careless? These are both much harder to answer for AI technology.

• “The UK is likely to resolve this problem for autonomous vehicles by introducing what is, in effect, strict liability on the insurers or owners of such vehicles.
In general

• “If the machine learning technology does not cause loss or damage directly, but instead assists humans in deciding how to act, then determining negligence liability becomes more difficult.

• “If the law were applied in this way, we are faced with the situation where the medical professional has probably discharged her duty of care because the technology is normally so accurate in its advice that it is reasonable to rely on it, while the producer of the technology owes no duty of care at all to patients. Thus, even if the patient’s loss were caused by carelessness on the part of the technology producer, there would still be no liability on the producer to compensate the patient.
Changes to law

• “In the short term, AI technology will be introduced gradually, and this gives time for lawmakers to appreciate the problem and consider ways to deal with it, but in the long term, the law of negligence will likely become an inappropriate mechanism for assigning responsibility and liability in the case of AI technologies which are in widespread use.

• “the courts need to be told how the AI made its decision. So, requiring transparency about the workings of AI might be a suitable interim solution to some of the legal problems, and has already been recommended as a tool for regulation.
US/UK

• Wide range of opinions
  • ACM Statement on Algorithm Transparency close to Europe
  • Strong neo-liberal, libertarian point of view
• Focus on immediate harm, not societal impact
• Competition
  • US has long history of exploiting technology and innovation
• Adaptive, evolutionary legal system as well as government regulation
  • Oliver Wendell Holmes, "The life of the law has not been logic; it has been experience."
China

A Next Generation Artificial Intelligence Development Plan

POSTED ON JULY 20, 2017  UPDATED ON AUGUST 1, 2017

This document was translated jointly by Graham Webster, Paul Triolo, Elisa Kania, and Rogier Cremers. John Costello assisted with helpful comments. An analysis of this document can be found on the New America website.

State Council Notice on the Issuance of the Next Generation Artificial Intelligence Development Plan

Completed: July 8, 2017

Released: July 20, 2017

A Next Generation Artificial Intelligence Development Plan

The rapid development of artificial intelligence (AI) will profoundly change human society and life and change the world. To seize the major strategic opportunity for the development of AI, to build China’s first-mover advantage in the
Next generation artificial intelligence development plan

- "AI has become a new focus of international competition.
- "AI has become a new engine of economic development.
- "AI brings new opportunities for social construction.
- "While vigorously developing AI, we must attach great importance to the potential safety risks and challenges, strengthen the forward-looking prevention and guidance on restraint, minimize risk, and ensure the safe, reliable, and controllable development of AI."
Kai-Fu Lee

- AI researcher/executive (Apple, Microsoft, Google), VC in China
- Book: *AI Superpowers, China, Silicon Valley, and the New World Order*
  - Argument that China will dominate AI and use this to overtake US
  - Europe does not play a role
AI is here and China will use it

• “For the past thirty years, Chinese leaders have practices a kind of techno-utilitarianism, leveraging technological upgrades to maximize broader social good while accepting that there will be downsides for certain individuals or industries.

• “But people in China are more accepting of having their faces, voices, and shopping choices captures and digitalized. This is another example of the broader Chinese willingness to trade some degrees of privacy for convenience.

• “When managing a country of 1.39 billion people — one in which 260,000 people die in car accidents each year — the Chinese mentality is that you can’t let the perfect be the enemy of the good. That is, rather than wait for flawless self-driving cars to arrive, Chinese leaders will likely look for ways to deploy more limited autonomous vehicles in controlled settings.

• “Google had taken six years to accumulate 1.5 million miles of real-world driving data. In just six months, Tesla had accumulated 47 million miles.
“China’s tech elites are aligned with the techno-optimistic American economists who believe that in the long run, technology always leads to more jobs and greater prosperity for all.

“For the past forty years, Chinese people have watched as their country’s technological progress acted as the rising tide that lifted all boats.

“Pervasive sense that the Chinese government will take care of all the displaced workers.”
“That vision of technology as a cure-all for global inequality has always been something of a wistful mirage, but in the age of AI it could turn into something far more dangerous.

“If left unchecked, AI will dramatically exacerbate inequality on both international and domestic levels. It will drive a wedge between the AI superpowers and the rest of the world, ...."
Blueprint for human coexistence with AI

• “We must proactively seize the opportunity that the material wealth of AI will grant us and use it to reconstruct our economies and rewrite our social contracts.

• “the most popular policy suggestions for adapting to the AI economy, many of them emanating from Silicon Valley. These three are largely “technical fixes,” tweaks to policy and business models that seek to smooth the transition but do not actually shift the culture.
• “The private sector is leading the AI revolution, and, in my mind, it must also take the lead in creating the new, more humanistic jobs that power it.

• “While AI handles the routine optimization tasks, human beings will bring the personal, creative, and compassionate touch."
China

- AI is a (strongly) government-supported endeavor
  - For its own (authoritarian) uses
  - For commercial reasons
  - For nationalist reasons
- Extremely positive view of benefits of technology progress
- Different expectation of privacy (?)
- Optimistic societal impact can be managed
Interconnected world...

- I don’t see anyway to reconcile these views
  - Tied to deeply rooted cultural values
  - Evolved over a long period
One Scenario

- Commercial pressures bring forth many (imperfect) innovations that use AI
- Popular demand / industry pressure precludes strong regulation in US and China
- Europe?
  - Pleas for restraint and ethics will have little weight in commercial race
  - Or, Europe sits out this round of innovation (again)

- Flaws and unintended consequences lead to regulation of most serious issues

- Necessity for open, public inquiry into causes
  - Airplane safety model
  - Trade limited liability for transparency

- Will this approach help with the social impact of AI?
Backup Slides
Airline Safety

Aviation Safety: Fatalities per trillion RPK (Revenue Passenger Kilometer)