Capgemini
RESEARCH INSTITUTE

Conversations

Quarterly review
N°3 — 2021

for tomorrow

Intelligent Industry:
The Next Era of
Transformation

#GetTheFutureYouWant

Conversations for tomorrow

PROFESSOR
MARINA JIROTKA
Professor of Human Centered Computing, Department of Computer Science, University of Oxford

RESPONSIBLE INNOVATION:
ESTABLISHING TRUST IN AUTONOMOUS SYSTEMS AND AI IN A CONNECTED WORLD

Professor Marina Jirotka leads a team focused on responsible innovation in a variety of advanced technologies and is at the forefront of recent work in Responsible Innovation in the UK and EU. She is director of the newly established Responsible Technology Institute at the University of Oxford and she is also an advisory board member of the Society for Computers and Law (SCL).

The Capgemini Research Institute spoke to Prof. Jirotka about the need to innovate responsibly; the repercussions that biased or untested autonomous systems can have for both organizations and individuals; and how academia, governments, and organizations can come together to establish standards for more ethical autonomous systems.

"

Not all the automated tools that organizations are considering for integration into their systems have been fully tested for ethical risk and lack of bias."

THE NEED FOR RESPONSIBLE AUTONOMOUS SYSTEMS

How relevant is responsibility in building autonomous systems and artificial intelligence (AI) today?

— Advances in engineering techniques and new technologies are driving the transformation of the manufacturing sector. New developments in machine learning, AI, and robotics have given rise to an expanding range of intelligent products, operations, and services. Moreover, this means that the algorithms that control these autonomous systems are now pervasive in society, to an extent to which people may be unaware. Autonomous vehicles, educational robots for children, and robots for old-age care interact with humans on a daily basis. Further, sector specific autonomous systems also support activities such as manufacturing, deep-sea mining, and space exploration programs.

Professor Marina Jirotka
Professor of Human Centered Computing, Department of Computer Science, University of Oxford

However, not all the automated tools that organizations are considering for integration into their systems have been fully tested for ethical risk and lack of bias. Users should be aware that design decisions taken while developing these intelligent products can have unintended consequences. For example, relying on historical data alone to train facial-recognition systems or applications for screening job applicants can lead to biases against certain groups of people.

Therefore, it is important to develop responsible approaches to building, monitoring, and evaluating technology to enhance individual autonomy and well-being by acting according to widely accepted human values.

What are the consequences for organizations of a misstep in matters of responsible innovation?

— Organizations today are acutely conscious of the need to be ethically aware and to be seen to be so under increased public and regulatory scrutiny. Many have created ethics boards to monitor their progress in meeting these new standards. However, even with such a board in place, it is challenging for management to maintain oversight and authority over the practices of the entire company. The problem of "ethics washing" (giving the appearance of commitment to ethical practices without rigorous implementing of the necessary steps) persists. Moreover, organizations will need to tackle the practical implementation of aspirational ethical charters in terms of designing algorithms to apply to autonomous systems. Unchecked datasets can entrench bias within a decision-making system.

The consequences of allowing the current situation of ethically under-monitored systems to continue could be grave. Whether it is a government department using off-the-shelf software, or a company that is developing an autonomous vehicle, a lack of care and effort in designing responsible and ethical systems runs the risk of seriously harming consumer and public trust. Put simply, if an organization today fails to recognize and align itself with the ethical standards demanded by its consumer base, then those same

consumers will take their business elsewhere. In contrast, if it considers its ethical position carefully and acts accordingly, it will not only establish itself as a brand that existing consumers can trust but will distinguish itself in a market where consumers are looking for more ethically sound options.

SYSTEMS THAT EVERYONE CAN UNDERSTAND

How important are explainability, transparency, and auditability in autonomous systems and AI? How can these systemic qualities be enhanced?

—The current lack of a uniform level of transparency across sectors makes it very difficult to establish an accountability structure to apportion responsibility for dealing with adverse incidents. Moreover, because most autonomous systems today are driven by generalized coding and algorithms, they lack a "human" perspective on how they will be used in the real world. An understanding of context is essential to effective implementation and cannot be gleaned from sifting through large datasets. Analysis of how previous systems have performed in real conditions is, therefore, a significant aspect of development.

"An understanding of context is essential to effective implementation and cannot be gleaned from sifting through large datasets. Analysis of how previous systems have performed in real conditions is, therefore, a significant aspect of development."

Within my own team as part of the RoboTIPS project, we are working closely with Bristol Robotics Lab to develop an "ethical black box." The concept is derived from that of an airplane's black box and the idea is to document the actions taken by a robot in the lead-up to an incident. Subsequently, the robot may be "interrogated" by accident investigators. Taken together with other evidence (such as recordings made by members of the public or time-stamping from sensors in the vicinity), investigators can arrive at a good understanding of the conditions that gave rise to the incident.

With new products, responsibility begins before the product even hits the drawing board. If, say, an autonomous vehicle has failed to recognize a stop sign because a sticker had been placed on the latter, that information could be shared with other developers without breaching the confidentiality of proprietary research. This is also where defined industry standards can be helpful; industries must collaborate and share data to minimize incidents and maximize equitable development which is part of the Trusted Autonomous Systems (TAS) network - RoAD (Responsible AV Data), investigating the ethical, legal, and societal challenges of using data from autonomous vehicles.

RoAD investigates the type of data that automated vehicles (AVs) should collect in order for their data recorder (black box) to be useful in any possible enquiry after an accident. Closely linked to the above, one of my students, Daniel Omeiza, is working on categorizing explanations and developing post-hoc explanation techniques for autonomous driving, working very closely with partners at the Oxford Robotics Institute.

Meaningful transparency is a fundamental principle of both good design and good business practice. Consumers should know how a system with which they are interacting has arrived at a certain decision that relates to them and could affect their lives. However, recognizing this imperative is just the first step; organizations must act to ensure that they are keeping up with developments in transparency.

A CLEAR WAY FORWARD FOR ORGANIZATIONS

What is your recommendation for organizations that are trying to build responsibility into their autonomous systems?

— With new products, responsibility begins at the moment of conception. The first step of responsible innovation is anticipating both the positive and negative consequences of a new product design or process. Rather than brushing the negatives under the carpet, these consequences should be brought into the light and examined from every angle, and strategies developed to prevent or mitigate them.

> **"The first step of responsible innovation is anticipating both the positive and negative consequences of a new product design or process."**

The idea of abiding by a code of "ethics by design" (the systematic inclusion of ethical principles in design systems and processes) is alluring; however, it is essential that developers grasp both the individual meaning of ethical practices and the overall ideology behind an organization's ethical framework in order to implement them effectively.

A practical approach is also required to helping developers to understand how to anticipate and deal with the problems that will emerge when systems are deployed in real-life contexts in which they will interact with humans.

The Responsible Technology Institute (RTI) at Oxford is working with the EPSCR's (Engineering and Physical Sciences Research Council's) "AREA" framework for responsible innovation.[1] According to this, a responsible innovation approach should be one that continuously seeks to:

- **Anticipate** – Describing and analyzing the impacts, intended or otherwise (for example economic, social, environmental), that might arise. This does not seek to predict but rather to support an exploration of possible impacts and implications that may otherwise remain uncovered and little discussed.

- **Reflect** – Reflecting on the purposes of, motivations for, and implications of the research, and the associated uncertainties, areas of ignorance, assumptions, framings, questions, dilemmas, and social transformations these may bring.

- **Engage** – Opening up such visions, impacts and questioning to broader deliberation, dialogue, engagement and debate in an inclusive way.

- **Act** – Using these processes to influence the direction and trajectory of the research and innovation process itself.

It is important to involve as many stakeholders as possible in the Engage stage – not only developers, but also civil society and members of the public. So for example while not everyone will have a clear understanding of how algorithms can impact their lives (and that different algorithms, used with different datasets, can have a range of consequences), creating this broader awareness is necessary, as is hearing a diverse range of opinions. Voices from different genders and cultures may help to ensure that systemic biases are addressed as the systems are being designed, rather than implemented in workarounds that are hastily brought in following a backlash.

The AREA framework is not a silver bullet, but it can be a good starting point for embedding responsible and ethical principles within an organization.

---

1 Engineering and Physical Sciences Research Council, "Anticipate, reflect, engage and act (AREA)".

How can enterprises, academia, and governments collaborate to ensure more responsible autonomous systems and AI are built and deployed?

— Often, the formal processes required to negotiate and effect change through governments and policymakers can move extremely slowly. There needs to be a much more agile pipeline between research, policy, government, and industry.

There has already been some positive change – organizations are now much more focused on societal challenges rather than profit alone and have been reaching out to academia to collaborate in finding effective solutions. The relationship between organizations and governments, however, is less advanced.

It is important to make this pipeline more flexible in terms of sharing information, processes, and tactics and strategies for mitigating incidents and issues, in order to ensure that responsible and ethical autonomous systems emerge going forward.

An emphasis on design that is ethical, unbiased, and transparent needs to be a priority for all parts of an organization, not just for developers. It is really important that this flows from the leadership to the rest of the organization; that it needs to be in all parts of the process – not just the design and the scoping, but from step one, from every person who has an impact on how products are built and what they are built for. And, in particular, that it not just be added on afterwards.

Professor Marina Jirotka
Professor of Human Centered
Computing, Department of Computer
Science, University of Oxford

"An emphasis on design that
is ethical, unbiased, and
transparent needs to be a priority
for all parts of an organization,
not just for developers."