For Reflection: Case Study Summaries

CASE 1: SUMMARY

Perpetuating unintentional scientific biases: When sampling is not intentionally diverse.
Even open-source data can contain discriminating biases, such as open-source code used for face recognition software.

THE REALITY: One in two adults in the US has their face in the facial recognition network, which is searched using algorithms that haven't been audited for accuracy.

THE CONSEQUENCE: We know our past is unequal, but to create a more equal future we have to look at the characteristics that we are optimizing for – for example, mistakenly recruiting for gender-diverse candidates using historically male-based strategies.

IN SUMMARY: Within the facial recognition community, benchmark data sets are meant to show the performance of various algorithms for easy comparison. The assumption is that if performance against a benchmark is good, then performance is good overall. Yet we haven't questioned the representativeness of the benchmarks, which may be providing a false notion of progress.

RECOMMENDED: Being transparent about the outcomes and about the bias tested for is one way of being accountable.

DISCUSSION

1. "Collecting data, particularly diverse data, is not an easy thing." BUT SHOULD IT BE A PRIORITY?
2. CONSIDER - "Wearing a white mask worked better than using my actual face."

   The implications are numerous, and profound:

   a. When used by the policeforce for facial recognition, technological bias is potentially converted into social bias
   b. Does poor capture of *real-world* diversity force our thinking through the narrow lens of what we then consider to be *possible* diversity?

CASE 2: SUMMARY

Technology is not neutral: The choices that get made in building technology then have social ramifications.
The medical profession has an ethic: First, do no harm. Are other research agendas missing the opportunity for bigger impact through a similar morality?

THE REALITY: Global accreditation for university science and engineering requires students have an understanding of related ethical issues.

THE CONSEQUENCE: Powerful tools like machine learning that can autonomously learn tasks by analyzing large amounts of data could ultimately alter human society — students should understand the potential consequences.

IN SUMMARY: "You can patch the software, but you can't patch a person" once damage is done.

RECOMMENDED: Educational institutions have a responsibility to play a leadership role in integrating these perspectives.

DISCUSSION

1. Should we be training the next generation of technologists and policymakers to consider the ramifications of innovations before products go on sale? Identifying issues that graduating students will have to grapple with in the next two, three, five, 10 years?
2. CONSIDER - The ethics and regulation of artificial intelligence:
   a. Autonomous weapons or self-driving cars
   b. Biased data sets including too few lower-income households to be representative of the general population
   c. Algorithmic risk scores used to forecast whether someone is likely to commit a crime, based on data such as whether a person was ever suspended from school, or how many of his or her friends have arrest records