Case Studies of Responsible AI/ML Failures

The following three case studies are all real-world examples of when AI has failed in some way. First, read through the following descriptions of each case:

1. Winterlight Labs auditory detection of Alzheimer's disease

   In 2016, Winterlight labs designed an AI-powered auditory test for Alzheimer's disease, where users' speech would be recorded and AI would be used to detect signs of Alzheimer's such as vocabulary richness, pauses in speech, and syntactic complexity. However, the initial research findings revealed a serious problem; non-native English speakers were being inaccurately flagged as having Alzheimer's disease. Since the data that was used to train the model had been collected from native English speakers from Ontario, Canada, this technology was unable to work reliably across different populations.

2. Wireless baby monitors hacked

   In 2018, there were several instances where wireless baby monitors were hacked which ultimately made national news headlines. In one case, a hacker used his newfound access to the baby monitor device to broadcast threats and shout sexual expletives. In another case, a more benevolent hacker used his newfound access to warn parents about the susceptibility of their device, in hopes that the parents would be able to address the situation before being targeted by nefarious hackers.

3. Hidden microphones in Nest devices

   In 2019, users of Nest Guard devices were shocked to discover hidden microphones inside the device. Users were concerned about the invasion of privacy as well as the breach of trust that resulted from not being properly informed about the specs of the device. From Google's perspective, the on-device microphone was simply a form of future-proofing that would allow the device to be compatible with updates that supported new functions later down the line.

Now, reflect on the following questions and join the discussion going on in the forum!

A. Which one of these cases do you find most concerning? Which concerns you the least?
B. What do you consider to be the relevant ethical challenges?
C. What do you think the designers of this technology could have done differently?
D. How can you apply learnings from these examples to your own job? Your personal life?
E. Do you agree or disagree with what others have already posted in the forum? Why?