Responsible AI Assignment 2 Ideas

Below is a list with suggested ideas for Assignment 2. You can choose one of these ideas, consider a topic that is proposed in the Assignment 1 Ideas document, or define your own idea.

1. Deep fakes detection

Case Description

Deep fakes are synthetically generated video, sound, or image content using different deep learning-based AI techniques such as GANs (Generative Adversarial Networks) and autoencoders. This implies training models with massive amounts of data that aim to convince humans that the resulted media content (i.e., the deep fake) is real. Currently, deep fakes are both a revolution and a critical threat in the social media domain as most of them are used for influencing people's beliefs and behaviors towards a topic, person, activity, process, or system in different societal domains.

However, deep fakes are also used for good purposes, e.g., in creative and marketing activities for enhancing the impact of current strategies and processes or building new ones. Thus, further developments in both generation and detection directions continue and show at a fast pace impressive result.

In this case, the focus is on deep fakes detection solutions. This implies analyzing either an existing detection solution (e.g., your own or found online using platforms such as GitHub or PapersWithCode) or a detection solution to-be-designed using materials such as the ones provided below.

Case Resources

For this case, you can read the state-of-the-art of deep fakes detection solutions provided in the articles by Masood et al. (2022) and Mirsky & Lee (2021).

- Masood, M., Nawaz, M., Malik, K. M., Javed, A., Irtaza, A., & Malik, H. (2022). Deepfakes Generation and Detection: State-of-the-art, open challenges, countermeasures, and way forward. *Applied Intelligence*, 1-53.
- Mirsky, Y., & Lee, W. (2021). The creation and detection of deepfakes: A survey. *ACM Computing Surveys (CSUR)*, *54*(1), 1-41.

Resources for finding existing solutions are below. Other platforms for finding such online solutions are accepted.

- https://github.com/topics/deepfake-detection
- https://paperswithcode.com/search?q_meta=&q_type=&q=deepfakes+detection

Case Description

2. Human oversight for drones

Human oversight for AWS (Autonomous Weapon Systems such as drones) means involving to a specific degree humans (a person or a group of people) in drone activities such as management, monitoring, and operations. This implies making sure the fact that AWS operate in a safe, secure, and responsible manner and their use is both legal and ethical, in other words, operations are conducted insider the boundaries of existing regulation, applicable laws, and respect social and ethical norms and values. Hence, human oversight is necessary for AWS to ensure that they operate and are used in a legal, ethical, and responsible way.

Case Description

To this end, both scientific experts and practitioners from different societal domains expressed their worries regarding diverse risks and potential ethical and legal issues surrounding the design, development, and deployment of AWS. Hence, in this assignment, you will either consider a particular case of AWS or how AWS should be designed using resources such as the ones provided below focusing on possible ethical issues. To this end, in both cases, a fundamental aspect that should be included is the human oversight levels and reasoning levels (i.e., technical, socio-technical, governance).

Case Resources

For this case, you can read and use the following resources:

- Steen, M., van Diggelen, J., Timan, T., & van der Stap, N. (2022). Meaningful human control of drones: exploring human–machine teaming, informed by four different ethical perspectives. *AI and Ethics*, 1-13.
- Verdiesen, I., Aler Tubella, A., & Dignum, V. (2021). Integrating comprehensive human oversight in drone deployment: a conceptual framework applied to the case of military surveillance drones. *Information*, *12*(9), 385.
- Stefik, M., Youngblood, M., Pirolli, P., Lebiere, C., Thomson, R., Price, R., ... & Schooler, J. (2021). Explaining autonomous drones: An XAI journey. *Applied AI Letters*, *2*(4), e54.

In this case, you can consider for instance systems such as Phoenix Ghost and Switchblade drones for which use descriptions can be found in the ongoing war in Ukraine.

3. Robot AI judge or lawyer

Robot AI judges are AI systems that are used to assess legal cases, e.g., law enforcement or military. This implies that such systems analyze the information received which represents evidence presented to them about the cases assessed. Based on the evidence provided, these systems execute a series of decision-making processes to determine appropriate outcomes. Just as in the case of human judges, robot AI judges process large amounts of data by merging internal information with the evidence provided which should produce corresponding, clear, unbiased, and transparent decisions to the legal cases analyzed. While systems proved to be efficient in a series of cases (still limited at this moment), it is expected that their use will continue to increase.

Case Description

AI-lawyers are AI systems that assist in legal cases (e.g., law enforcement of military) based on the practice of law. This implies automated legal (routine) activities such as document review and legal research plus legal advice, recommendations, and decisions. While these systems proved to be efficient and help reducing the time allocated for legal assessment, currently, such systems have an advisory role by augmenting human legal capabilities of lawyers and increase their efficiency and overall productivity.

In this assignment, you will choose either a robot AI judge or AI lawyer to investigate/proposed its design.

Case Resources

For this case, you can read and use the following resources:

- Campbell, R. W. (2020). Artificial intelligence in the courtroom: The delivery of justice in the age of machine learning. *Colo. Tech. LJ*, *18*, 323.
- Sourdin, T. (2018). Judge v Robot?: Artificial intelligence and judicial decision-making. *University of New South Wales Law Journal, The*, *41*(4), 1114-1133.
- Example of AI judge described in https://www.thestatesman.com/opinion/robot-judges-chair-1503031697.html
- Markovic, M. (2019). Rise of the robot lawyers. *Ariz. L. Rev.*, *61*, 325.
- Xu, N., Wang, K. J., & Lin, C. Y. (2022). Technology Acceptance Model for Lawyer Robots with AI: A Quantitative Survey. *International Journal of Social Robotics*, *14*(4), 1043-1055.
- Example of AI lawyer: DoNotPay described at https://eu.usatoday.com/story/tech/2023/01/09/first-ai-robot-lawyer-donotpay/11018060002/ and https://www.npr.org/2023/01/25/1151435033/a-robot-was-scheduled-to-argue-in-court-then-came-the-jail-threats

4. Fatigue Detection System

Case Description

A common problem in road safety is that drivers are too tired or too distracted to drive. Businesses that depend on the transport of people, goods and materials, such as oil and gas companies, taxi services and bus companies, therefore, look for options to detect fatigue and distraction in their drivers.

A common approach is to install a camera that analyses the face of the driver. If the eyelids of the driver close, this is a sign that the driver may be too tired to drive. Likewise, when the driver looks inside the cabine, they may be too distracted to pay attention to the road.

Case Resources

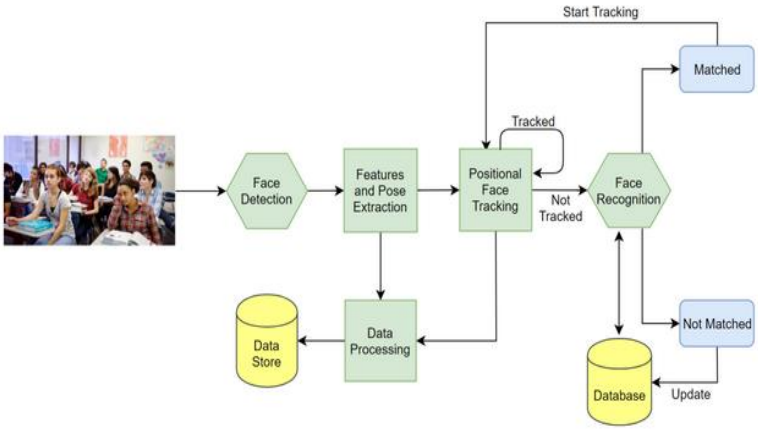
You can use one of the systems from the overview of fatigue detection systems for this case study: the Nauto system. For this, visit the following website: https://www.nauto.com/ and collect information about how the system functions. Furthermore, read online reviews of this system and find out what drivers indicate as possible downsides of the system. To this end, you can consider downsides and ethical issues of this system that could be tackled in this assignment.

5. Monitoring student attention

Case Description

As a lecturer you want to make sure that your students pay attention to your class and understand the materials that you are presenting. In a large class, however, it can be difficult to monitor every students' attention while also presenting the materials. This becomes even more difficult in an online setting, where not every student may be visible on their webcams.

Canedo and colleagues therefore propose a computer vision that helps lecturers with monitoring their students' attention.



Case Resources

In your analysis, you can take into consideration how the system works, what is being recorded, how it is analyzed, what is being predicted, what is done with the predictions etc. and connect these aspects with possible ethical issues.