The Dangers of Drowsiness Detection: Differential Performance, Downstream Impact, and Misuses
Jakub Grzelak and Martim Brandão, AIES 2021

KING'S College LONDON

Motivation
- Drowsiness detection (DD) algorithms **promise to improve driving/work safety and efficiency** by estimating state of alert from images or other data; triggering warnings or "supervisor" intervention.
- Use of such algorithms could lead to **downstream disparate impact and malpractices.**

Audit of differential performance in DD algo. [1]

	Alert	Low vigilant	Drowsy
Alert	0.63	0	0
Low vigilant	0.37	0.12	0.25
Drowsy	0	0.88	0.75

Confusion matrix for "Middle Eastern" group

Group	VA	VA (out-of-dist)
Male	57.5%	56.2%
Female	70.4%	81.5%
Caucasian	66.7%	60.0%
Non-white Hispanic	66.7%	66.7%
Indo-Aryan & Dravidian	56.7%	58.9%
Middle Eastern	50.0%	62.5%
East Asian	66.7%	47.6%

Per-group visual accuracy

Consistent results:
- Lower accuracy on Male, Indo-Aryan & Dravidian, Middle Eastern.
- Especially harmful to Middle Eastern and Indo-Aryan & Dravidian (biased towards drowsy).
- Lack of generalization to new groups.

Potential downstream impact due to biased performance
Some groups more likely to be wrongly considered drowsy.
- Scenario 1: Ignoring some driver input if drowsy?
- Scenario 2: Preventing to turn engine on if drowsy?
- Scenario 3: Sounds and physical stimuli to "awake" driver?
In all cases → **higher risk, driver stress/annoyance, harm, consistently for these groups.**
- Scenario 4: Usage-Based Insurance with drowsiness signals
→ **higher premiums for already marginalized groups.**

Academically-promoted vs real use
- Academic research promotes DD for driver/worker/student monitoring for "safety" and "efficiency"
- Narrative does not match **real interests and uses driven by profit: insurance and worker management/control.**

Potential misuses
- **Insurance misuses:** coverage avoidance, cream skimming
- **Worker surveillance and control:** increase control over truck drivers for higher throughput (fewer rest time), challenging drivers' claims of tiredness, driver performance rankings...

References
[1] Ghoddoosian, et al. 2019. "A realistic dataset and baseline temporal model for early drowsiness detection", CVPR 2019 Workshops.