DigitalWorld 2023 & NexComm 2023

Theme

Cohabitation of Humanized AI and Humans for Critical Systems Resilience

PANEL: AI Hour!

VENICE
April 2023

Passing the torch of Control both senses; one of the major dilemma
Threshold of passing the responsibility.

On human decisions: stress, common sense; Human panic, Hesitation to act, Emotions

Avalanche of alarms; correlation is needed, to have only one 'representative' situations

Legal implications, liability

An 'apriori recognizable threshold to act' and routine control

Security issues; human as intruder
Good/trustees: only skilled /certified decision humans
Preparedness: situational training

CONTRIBUTORS

Moderator

Prof. Dr. Lasse Berntzen, University of South-Eastern Norway, Norway

Panelists

Prof. Dr. Oliver Michler, TU-Dresden, Germany

Nadine-Christine Wessel, Technische Hochschule Brandenburg, Germany

Prof. Dr. Petre Dini, IARIA USA/EU

Chair Position

VENICE
April 2023

- Discussing key issues and challenges that may arise when integrating AI and human decision-making systems
- When to use AI?
  - Sometimes AI can handle simple tasks to offload human work, but AI may be biased
    - AI may also cause deskilling
    - Example: Recruiting
  - At other times AI can intervene to reduce or avoid errors made by humans (stress, panic, emotions)
    - Example: Collision avoidance

Chair Position

VENICE
April 2023

- There are also legal and liability considerations associated with integrating AI and human decision-making, particularly in critical systems where the consequences of errors or failures can be significant
- Also, ethical considerations
  - Autonomous cars and critical decisions

Chair Position

VENICE
April 2023

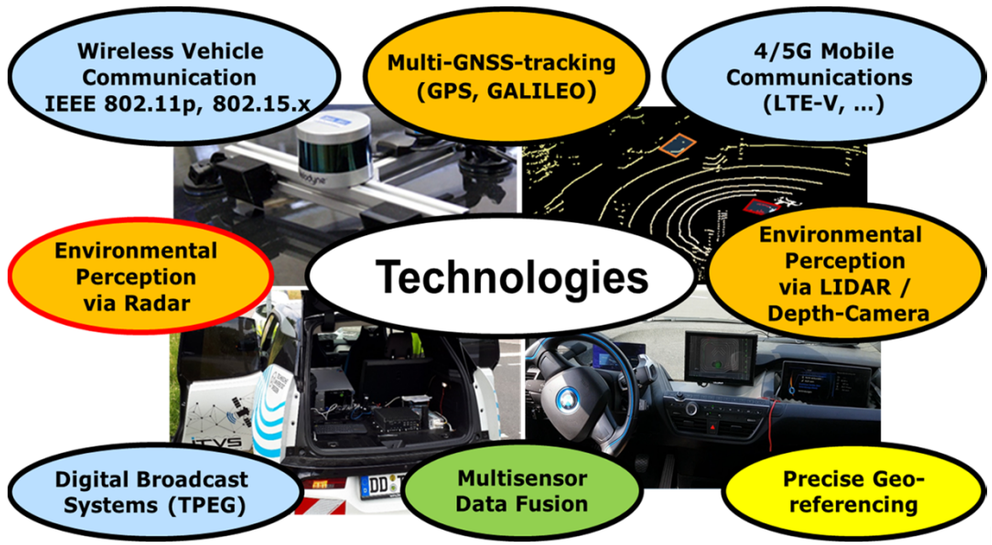
- In addition, there are security concerns to consider when integrating AI and human decision-making, as there may be risks associated with human intruders or other threats
- This may require limiting decision-making authority to only skilled and certified individuals and providing situational training to ensure preparedness for potential scenarios

Panelist Position

VENICE
April 2023

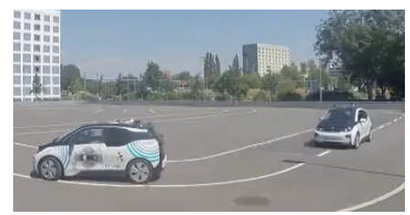
Autonomous driving and Critical system structures with AI - Coexistence and Synergy in technological-algorithmic Processes

- **Technological aspects …**
- **Challenges of AI …**
  - AI advantages: simple, fast, robust for learned processes, energy efficient
  - AI disadvantages: no determinism, reproducibility, open-ended for untrained scenarios
- **Synergetic role of AI …**
  - Integrety: AI as underlaying system as well as human knowledge overlaying safeguarding enveloping system
- **Legal implications, liability for accidents …**
  - Question of responsibility (manufacturer, OEM, SW code)
  - Decision basis (self-protection, material damage, personal damage)

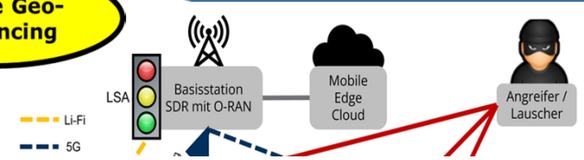


Human (Det/Stat) - based Knowledge
(Securing, Confidentiality, Integrity)

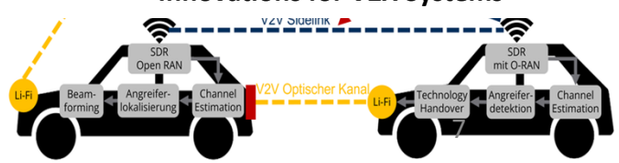
AI/ML – based Knowledge
(State detection, Controls, …)



https://www.youtube.com/watch?app=desktop&v=2kgUBZe4E5c



Project PHYSICS – Physical Layer Security Innovations for V2X Systems



Panelist Position

VENICE
April 2023

On the Unintended Consequences of the Interactions between Humans and Intelligent Decision Support Aids: Managing the Threat of Decisional Deskilling

Nadine-Christine Wessel

- Deploying (partially) intelligent Decision-Support Systems (iDSS) in the workplace can lead to **unintended organizational consequences**
  - Employee-related: loss of critical thinking, knowledge (both declarative and procedural), expertise
  - IT-Compliance: misuse of systems. system circumvention
- Interacting factors can shape deskilling effects
  - Individual factors: users' knowledge and skills, attitudes, and motivations toward using iDSS.
  - Technical factors: design and functionality -> ease of use, availability of relevant information
  - Organizational factors: policies & procedures governing the use of iDSS, level of support provided to users
- **Decisional Deskilling may remain hidden until iDSSs are disrupted or discontinued**, even though it can occur on a latent level
- Contributing factors include the degree to which iDSSs take over decision making activities, dependence on iDSS (reliance), and time spent with iDSS (?)
- Mitigation practices for Decisional Deskilling include training and support programs, monitoring employee reliance, **increasing participation in human decision making**, and re-evaluating iDSSs effectiveness.

Panelist Position

VENICE
April 2023

- **Assisting: Decision paradigm change**
  AI and the Human brain need to be used in conjunction; one is not to replace the others; *bias* and *trade-offs* must be cleared by humans.
  Human + AI based decisions need *personalized H-AI context-based training*.
  The feedback loop must be updated based in *success-index* of cooperation (see Intelligent Tutorial Systems).
  Revisit: Replace - Assist - Advice | Feedback-by-request, Suggestions-by-predictions
  Validation: Validate synchronization duration | Validate use case scenarios
  Flow: AI (Personalized-AI) & in-Context (Human skills) -->> Assisted decisions

- **Personalized AI for critical system resilience**
  Metaverse is training the AI with virtual models;
  Simulations accelerate the system deployments but no real-world environments.
  Latency, reliability, synchronization (for collaborative work)

  Digital-twin approach can be used for a dry-training AI-Humans for specific tasks
  Decision prediction and *feedback as options*.
  Updated training and *real-time impact feedback*.

OPEN DISCUSSION

VENICE
April 223

Output highlights

- The decisions should be taken by humans; AI only assists.
- Feedback Human-AI must be bidirectional; no AI decision should be enforced by an AI entity
- Evaluation of easy/hard decisions (when time-related situation, AI-based tools might excel; in some situations, AI- is even better than a human, question of human skills and training)
- AI-based training (for skills updates; e.g., avionics)
- Great risk of deskilling by ‘lose responsibility’ or lack of responsible decisional actions
- Legal responsibilities should be well-defined (e.g., failed sensors might trigger wrong decisions when on ‘automatic control)
- The risks need to be controlled by special certifications and role definitions